UMN-TH-916/90

Quantum Geometrodynamics*

M.J. Duncan[†]

Department of Physics, University of Minnesota, Minneapolis MN 55455

There is a theory which states that if ever anyone discovers exactly what the Universe is for and why it is here, it will instantly disappear and be replaced by something even more bizarre and inexplicable.

There is another theory which states that this has already happened.

D. Adams

*Lectures presented at the TASI summer school, Boulder, Colorado, June 1990. To be published in the proceedings of the school, editors P.Langacker and M.Cvetic.

[†]E-mail: duncan@vx.acs.umn.edu

1 Introduction

The past few years have seen considerable activity in the fields of quantum cosmology and wormholes. It is becoming increasingly clear that quantum cosmology is a reliable effective theory for cosmology in the early stages of the universe at which classical physics breaks down. Also clear is that the wormhole program can yield real testable predictions concerning the values of some fundamental constants of nature. Both these subjects may be lumped under the general heading of Quantum Geometrodynamics (QGD), a term coined by Wheeler some thirty years ago. In these introductory lectures to the subject I shall be discussing the basic elements of quantum cosmology and its applications. I shall also talk about the elements of the wormhole program and where they are headed. Fuller discussions of the foundations of wormhole physics can be found in previous TASI lecture series [1].

Whatever the successful theory incorporating quantum gravity may be, it should contain within it the unique requirements for reproducing the universe as we know it today. At present we have no inkling as to what this theory might tell us about cosmology, and so we have to rely on what we do know and extrapolate towards the Planck scale. If the action for general relativity is still valid as we approach the Planck scale, before quantum fluctuations make it inappropriate, we could be justified in building from it an effective quantum theory for the long wavelength modes of gravity. In particular, this description may be the appropriate one for discussion of cosmology. In a sense we may regard the degrees of freedom of the cosmology as the remnants of quantum gravity that ultimately give rise to the large scale structure of the universe. It is then a short step to consider the effective quantum theory as being formulated in terms of the cosmological degrees of freedom.

The clearest route to quantum cosmology is through the Hamiltonian formulation, as described in the next section. When we quantize, the Hamiltonian becomes an operator which acts on a wave function. The wave function of the universe is a strange object, as you might expect, and it contains many possible histories of spacetime. We shall discuss how classical cosmologies naturally emerge from the formalism in obeisance to the correspondence principle. Even though the wave function contains many cosmologies, in certain circumstances it will tend to prefer one over the rest. In one of the applications of QGD the initial conditions

for inflation can be extracted. The selection of one universe from a set of possibilities may sound metaphysical, but all that is happening is that the universe is selecting for itself its final form.

Even though the universe's geometry may have a set of possible histories, you might seek refuge in the knowledge that the laws of nature are immutable. However, even this is uncertain when we consider the full implication of the indeterminacy of geometry. Wormholes are tunnelling configurations, instantons, between geometries and their effect is to render the fundamental constants of nature dependent on parameters external to our universe. It would appear that metaphysics has crept in by the back door, but the startling result is that there *are* preferred values. For example, we can show that the preferred value of the cosmological constant is extremely small, an observational fact that has never had a satisfactory explanation. I shall also discuss the third quantization of the wave function of the universe, as it is an alternative method for obtaining the same result.

For further reading on the basics of quantum cosmology I can recommend an old set of lectures by Hawking [2]. There is also published a bibliography of papers on the subject [3], which is helpful for learning about its applications. Finally, there are the standard textbooks [4] which may be consulted if some of the general relativity is unfamiliar[1].

2 The Hamiltonian formulation of gravity

In quantum cosmology it is convenient to reformulate gravitation as a Hamiltonian system. The Schrödinger equation of the more familiar quantum mechanics will have an analogue when we quantize the gravitational degrees of freedom. The total action for matter coupled to gravity in a spacetime with a Lorentzian signature is,

$$S_{tot} = \frac{1}{16\pi G_N}\int_{\mathcal{M}} (R - 2\Lambda)\sqrt{-g}\, d^4x + S_{\partial\mathcal{M}} + \int_{\mathcal{M}} \mathcal{L}_\phi \sqrt{-g}\, d^4x \qquad (1)$$

This is the form of the effective action at energies much lower than the Planck scale $M_P = G_N^{-1/2}$ where quantum fluctuations in the spacetime are expected to invalidate (1). The spacetime manifold (de Sitter space for example) is denoted $\mathcal{M}$ and it may have a boundary $\partial\mathcal{M}$. If there is a boundary we must add on the correction $S_{\partial\mathcal{M}}$, more about which later.

[1] Throughout the lectures I shall be using the conventions of Misner, Thorn and Wheeler.

We are generically denoting the matter fields by ϕ with Lagrangian density $\mathcal{L}_\phi$. We have also allowed for a cosmological constant Λ. As usual, variation of this action with respect to the independent matter degrees of freedom will yield their equations of motion. Similarly, variation of the action (1) with respect to the metric $g_{\mu\nu}$ will give the Einstein equations,

$$G_{\mu\nu} = \left(R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R\right) = 8\pi G_N T_{\mu\nu} - \Lambda g_{\mu\nu} \qquad (2)$$

Given a matter distribution with stress energy tensor $T_{\mu\nu}$ and a cosmological constant Λ we can, in principle, solve the Einstein equations for the geometry.

The action is invariant under general coordinate transformations. The whole point of general relativity is that coordinate systems are irrelevant. Thus we should expect that not all ten degrees of freedom in the metric are physically independent, some of them really reflect the reparametrization (gauge) invariance of the theory. To separate the dynamical degrees of freedom from these gauge degrees of freedom it is convenient to write the metric in the ADM form [5]. We imagine slicing the spacetime up into a series of spacelike hypersurfaces. Each of these three-dimensional hypersurfaces has its own three-dimensional geometry with its own three-metric h_{ij}. The metric of the spacetime may be written,

$$ds^2 = g_{\mu\nu}\, dx^\mu \otimes dx^\nu = -\left(N^2 - N_i N^i\right) dt \otimes dt + 2N_i\, dt \otimes dx^i + h_{ij}\, dx^i \otimes dx^j \qquad (3)$$

The N and N_i are called the lapse and shift functions respectively. They are so named because different choices correspond to different parametrizations of the time coordinate, and different embeddings of the three-spaces into the spacetime. They are just our gauge degrees of freedom.

The spacelike slices will have their own internal geometry with metric h_{ij} from which we can derive the *intrinsic* curvature of the three-space. This is denoted ${}^3R^i{}_{jkl}$ and is derived from the three-metric in the usual manner. Another important quantity is the *extrinsic* curvature which arises as follows. Each slice has a normal vector field, whose value over the slice is dependent on the way the slice is embedded in the four-dimensional spacetime. The way in which the normal varies across the slice is described by the extrinsic curvature, which is defined on the three-space itself. For the ADM metric (3) it has the form,

$$K_{ij} = \frac{1}{2N}\left[\nabla_i N_j + \nabla_j N_i - \dot{h}_{ij}\right] \qquad (4)$$

where ∇_i is the covariant derivative on the slice, derived from the three-metric. The surface term $S_{\partial\mathcal{M}}$ ~~alluded to above~~ is written in terms of the three-metric and extrinsic curvature of the slice that constitutes the boundary to the spacetime [6],

$$S_{\partial\mathcal{M}} = \frac{1}{8\pi G_N}\int_{\partial\mathcal{M}} K\sqrt{h}\,d^3x \tag{5}$$

where the trace of the extrinsic curvature is $K = h^{ij}K_{ij}$, and h^{ij} is the inverse of the matrix h_{ij}. The reason we introduce this correction is to remove second time derivatives from the action (1) upon integration by parts (this ensures that the action from $\partial\mathcal{M}_1$ to $\partial\mathcal{M}_2$ plus the action from $\partial\mathcal{M}_2$ to $\partial\mathcal{M}_3$ equals the action from $\partial\mathcal{M}_1$ to $\partial\mathcal{M}_3$). Upon doing so the action takes the form,

$$S_{tot} = \frac{1}{16\pi G_N}\int_{\mathcal{M}} N\sqrt{h}\left[K^{ij}K_{ij} - K^2 + {}^3R - 2\Lambda\right]d^4x + \int_{\mathcal{M}} N\sqrt{h}\,\mathcal{L}_\phi\, d^4x \tag{6}$$

where the lowering and raising of the indices is done with the three-metric h_{ij} and its inverse h^{ij}. The boundaries to the spacetime can be thought of as the initial and final end points of the system. We can find the equations of motion by varying this action with respect to the variables, while keeping them fixed at the end points. Given the three-metric and the matter fields on both initial and final hypersurfaces we could solve the equations of motion to find the four-geometry that interpolates between the two slices and which extremizes (6).

It is immediately clear that (6) contains no time derivatives of the lapse and shift functions, reflecting their rôle as gauge rather than dynamical degrees of freedom. Variation of the action with respect to them will provide constraints. The independent dynamical degrees of freedom are $h_{ij}(\mathbf{x})$ and $\phi(\mathbf{x})$, the three-metric and matter fields on a slice *i.e.* one degree of freedom at each point $\mathbf{x}$ on the hypersurfaces. Variation with respect to the three-metric will give the Euler-Lagrange equations for the geometry, and variation with respect to the matter fields will give their equations of motion. The Euler-Lagrange equations are equivalent to the Einstein equations (2). Solving them will give the time developments or history $h_{ij}(\mathbf{x},t)$ and $\phi(\mathbf{x},t)$. As stated above, we could fix the boundary conditions on the solutions, to find the interpolating spacetime, by specifying the metric and field configurations on two hypersurfaces. Alternatively we could specify initial data on only one slice and use the equations to determine the geometry and matter field configurations induced on any other. The initial data in this case are the three-metric and matter fields on the particular hypersurface along

with their first time derivatives. However, the time derivative of the metric is not quite a fundamental geometrical object on the slices *per se*. The Hamiltonian formulation remedies this.

It is a straightforward task to obtain the Hamiltonian description for the system. From the action (6) and the dependence of the extrinsic curvature on the three-metric from (4) we can identify the momenta conjugate to the h_{ij} as,

$$\pi^{ij} = \mu\sqrt{h}\left(h^{ij}K - K^{ij}\right) \tag{7}$$

where we have written $\mu = M_P^2/16\pi$. From the matter action we can obtain π_ϕ the momenta conjugate to the fields as usual. The momenta (7) depend only on the three-metric and the extrinsic curvature, two truly geometrical objects defined on each slice. The action can now be re-expressed in the canonical form,

$$S_{tot} = \int_{\mathcal{M}}\left[\pi^{ij}\dot{h}_{ij} + \sum_\phi \pi_\phi\dot{\phi} - NH - N_i\chi^i\right]d^4x \tag{8}$$

The Hamiltonian is a function of the dynamical degrees of freedom and their conjugate momenta. When expressed in these terms it can be readily seen to have the form,

$$H = \frac{1}{2\mu}G_{ij,kl}\pi^{ij}\pi^{kl} + \mu\sqrt{h}\left(2\Lambda - {}^3R\right) + H_\phi \tag{9}$$

The $G_{ij,kl}$ is symmetric under interchange of the first pair of indices with the second pair and is simply a function of the three-metric,

$$G_{ij,kl} = \frac{1}{\sqrt{h}}\left(h_{ik}h_{jl} + h_{il}h_{jk} - h_{ij}h_{kl}\right) \tag{10}$$

The H_ϕ is the part of the Hamiltonian due to the matter fields. It is related to the energy of the matter fields, $H_\phi = \sqrt{h}\bar{H}_\phi$, where $\bar{H}_\phi$ is the energy density. This is the usual normal component of the stress energy tensor $\bar{H}_\phi = n_\mu n_\nu T^{\mu\nu} = N^2T^{00}$ as the normal vector field for the ADM metric is $n_\mu = (-N,0,0,0)$. There is a similar contribution to the χ^i which has the form, $\chi^i = -2\nabla_j\pi^{ij} + \chi^i_\phi$. As an example we give them for a real scalar field;

$$\begin{aligned} \bar{H}_\phi &= \frac{1}{2h}\pi_\phi^2 + \frac{1}{2}(\nabla\phi)^2 + V(\phi) \\ \chi^i_\phi &= \pi_\phi\nabla^i\phi \end{aligned}$$

It should be clear from (8) that the constraints which derive from variation of the action with respect to the lapse and shift functions are that both H and χ^i vanish. They reflect the fact that there is no absolute meaning to 'time' and no preferred embedding of the spacelike hypersurfaces in general relativity. The same geometry can be described by many different choices of coordinate systems t and $\mathbf{x}$. The vanishing of the Hamiltonian can also be thought of as just saying that the sum of gravitational and matter energy in the system is zero – they balance each other. Moreover, the constraints imply that the variables and their momenta cannot have arbitrary initial conditions on some slice. This might seem a little strange on first encounter. In classical mechanics the Hamiltonian (if time independent) is a first integral of the equations of motion and is equal to an integration constant, the total energy of the system, determined by the initial conditions. Another reason that the Hamiltonian here is fixed for us is that the energy appears directly in the equations of motion (2) and not as a first integral. The source of this is the principle of equivalence. In classical mechanics it is the concept of force which is central whereas energy is secondary. There the forces acting on a particle result in a deviation from inertial motion. In general relativity, however, the principle of equivalence says that there is no such thing as a gravitational force and no way to construct inertial frames. Natural motion is on a geodesic, determined by the spacetime, and forces don't enter into the picture at all. As we shall see, the constraint $H = 0$ lies at the heart of quantum cosmology. The constraint $\chi^i = 0$ is just a consequence of the internal three-dimensional coordinate invariance of the slices, meaning that it is truly the three-geometry that is dynamical and not any particular parametrization of the three-metric.

3 Quantum Cosmology

We have started with the action for general relativity and recast it as a Hamiltonian system in order to separate the dynamical from the gauge degrees of freedom. We have found that the system is constrained, meaning that we cannot specify both the variables and their momenta arbitrarily on any given slice of the spacetime. One of the constraints is that the Hamiltonian vanishes, $H(\pi^{ij}, h_{ij}; \pi_\phi, \phi) = 0$. The transition from a classical to a quantum system is quite straightforward. Starting from the hamiltonian formulation of gravity we postulate canonical

quantization rules for the degrees of freedom and their conjugate momenta. Our rationale for this is that whatever quantum gravity might be, it certainly ought to reproduce cosmology in some limit. Quantum cosmology is simply an effective theory for quantum gravity at energies below the Planck scale. We should not expect our action (1) to be valid anywhere close to the Planck scale. The quantized degrees of freedom of the cosmology are the remnants of the degrees of freedom of the complete quantum theory of gravity that give rise to the large scale structure of the universe. We therefore make the operator substitution $\pi^{ij}(\mathbf{x}) = -i\delta/\delta h_{ij}(\mathbf{x})$ and $\pi_\phi(\mathbf{x}) = -i\delta/\delta\phi(\mathbf{x})$. The Hamiltonian constraint becomes the Wheeler-DeWitt (WDW) equation [7],

$$\mathcal{H}\left(-i\frac{\delta}{\delta h_{ij}}, h_{ij}; -i\frac{\delta}{\delta\phi}, \phi\right)\Psi(h_{ij};\phi) = 0 \tag{11}$$

which is the gravitational equivalent of the Schrödinger equation. The Ψ is a functional over the space of all three-metric configurations and matter configurations on a slice. This infinite-dimensional functional space is commonly called *superspace* (this subject has been around longer than supersymmetry) and the Ψ has been called 'the wave function of the universe' and 'the wave function of everything'. As you may have already surmised, the WDW equation has some peculiarities.

The most apparent is that it seems to refer to only one slice. The Schrödinger equation has a time derivative of the wave function on its right hand-side, whereas it is zero here. This wave function has no history in the conventional sense. This is a consequence of the gauge invariance from the lapse function which says that the 'time' between any two slices is meaningless. Because of this, our conventional Copenhagen interpretation of a wave function is doomed from the outset. It is also doomed because there is no 'observer' larger than the universe. We shall postpone a discussion of this until we have developed the formalism a bit further. A closer look at (11) reveals that it resembles more the Klein-Gordon (KG) equation rather than the Schrödinger equation. The WDW equation is hyperbolic, as the superspace metric $G_{ij,kl}$ has signature $(-+++++)$, the minus sign being associated with the scale $\sqrt{h}$ of the three-metric. This will have consequences for our interpetation, as we shall see later.

It is possible to formally develop quantum cosmology for the general case above. The problem, though, is that solving infinite-dimensional functional differential equations is not an easy task and is a branch of mathematics still in its infancy. As we are interested in

cosmology we are fortunate that we rarely need the complete $h_{ij}(\mathbf{x})$. For example, the assumption of homogeneity immediately reduces the three-metric to a function of *only* a finite number of variables as the three-geometry is the same no matter where we are on the slice. In fact, the most commonly used cosmological models are the Friedmann Robertson Walker (FRW) types which have a metric of the form,

$$ds^2 = -dt^2 + a^2(t)\left[\frac{dr^2}{1-kr^2} + r^2 d\Omega\right] \tag{12}$$

where $k = -1, 0, 1$ for open, flat and closed three-spaces respectively. These cosmologies are isotropic in addition to the homogeneity. As you can see, there is only *one* degree of freedom in this metric and it is the scale factor a. We could just as well have started with this metric and arrived at the corresponding WDW equation, now a differential equation of a finite number of variables – something we at least have a fighting chance of solving. The truncation of the metric to a finite number of degrees of freedom is called the *minisuperspace* approximation in quantum cosmology. Practically all quantum cosmology is studied in minisuperspace, and we shall do so too for the sake of clarity.

Let us begin with the homogeneous, but not necessarily isotropic, closed cosmology known as Bianchi IX to see how this works. The metric for this can be written as,

$$ds^2 = \left(\frac{2G_N}{3\pi}\right)\left(-N^2(t)\,\mathrm{d}t\otimes\mathrm{d}t + a^2(t)\left[e^{2\beta(t)}\right]_{ij}\boldsymbol{\sigma}^i\otimes\boldsymbol{\sigma}^j\right) \tag{13}$$

which requires a little explanation. Due to the homogeneity the degrees of freedom can only be time dependent. The N is the lapse function as usual, and we have ignored the shift functions as we have chosen to write the three-metric in a particular parametric form. The overall factor out front is just a convenient scaling as we shall see in a moment. The three-space is closed and has the geometry of a distorted three-sphere. The best way to describe this geometry is to construct a homogeneous basis for a unit three-sphere $\boldsymbol{\sigma}^i$. The scale factor for the space is a while the traceless matrix β_{ij} parametrizes the anisotropy. An observer on the three-space would not be able to tell where they were (homogeneity) but they would be able to tell in which direction they may be facing (anisotropy). The basis of one-forms for a three-sphere satisfies the structure equation,

$$\mathrm{d}\boldsymbol{\sigma}^i = -\epsilon_{ijk}\,\boldsymbol{\sigma}^j\wedge\boldsymbol{\sigma}^k \tag{14}$$

where ϵ_{ijk} is the familiar permutation symbol, and a convenient representation is in terms of the Euler angles of the rotation group,

$$\begin{aligned}\boldsymbol{\sigma}^1 &= \frac{1}{2}(\sin\psi\,\mathrm{d}\theta - \cos\psi\sin\theta\,\mathrm{d}\phi)\\ \boldsymbol{\sigma}^2 &= \frac{1}{2}(\cos\psi\,\mathrm{d}\theta + \sin\psi\sin\theta\,\mathrm{d}\phi)\\ \boldsymbol{\sigma}^3 &= \frac{1}{2}(\mathrm{d}\psi + \cos\theta\,\mathrm{d}\phi)\end{aligned} \tag{15}$$

Finally, the traceless matrix β_{ij} may be chosen to be diagonal in which case it has only two independent components. We may write $\beta_{ij} = \mathrm{diag}(\beta_+ + \sqrt{3}\beta_-, \beta_+ - \sqrt{3}\beta_-, -2\beta_+)$. The picture we have from (13) for the slices is a distorted three-sphere with volume proportional to a^3 (since $\det(e^{2\beta}) = 1$) and where the $\beta_\pm$ parametrize the ratios of lengths in different directions. In the case where the $\beta_\pm$ vanish the space is isotropic and you can see for yourself that the Bianchi-IX metric reduces to the closed FRW metric (12) as expected. More detailed discussions of homogeneous spaces and Bianchi models may be found in the textbooks [4].

From this metric (13) we may work out the curvature scalar and then the action (1) for this space[2]. The gravitational part of the action is simply,

$$S_{\mathrm{IX}} = \frac{1}{2}\int dt\left[-\frac{a\dot{a}^2}{N} + \frac{a^3}{N}\left(\dot{\beta}_+^2 + \dot{\beta}_-^2\right) - \lambda N a^3 + \frac{a^3}{6}N\,{}^3\mathcal{R}_{\mathrm{IX}}\right] \tag{16}$$

which should be compared with (6). The overall factor in (13) was included to remove the $16\pi G_N$ and make everything expressible in Planck units. As the space is homogeneous the degrees of freedom only depend on the parameter t and so the integral over the three-space can be performed. In this case $\int\sqrt{h}d^3x = \left(\frac{2G_N}{3\pi}\right)^2(2\pi^2)a^3$ as the volume of a unit three-sphere is $\int\boldsymbol{\sigma}^1\wedge\boldsymbol{\sigma}^2\wedge\boldsymbol{\sigma}^3 = 2\pi^2$. We have also chosen to scale the cosmological constant according to $\Lambda = \left(\frac{9\pi}{2G_N}\right)\lambda$, to make the action a function of dimensionless parameters. The intrinsic curvature for the Bianchi-IX space is,

$${}^3\mathcal{R}_{\mathrm{IX}} = \frac{2}{a^2}\mathrm{Tr}\left(2e^{-2\beta} - e^{4\beta}\right) \tag{17}$$

In the special case of isotropy it reduces to $6/a^2$, the curvature of a three-sphere of radius a. The gravitational part of the Hamiltonian is found by variation with respect to the lapse

[2] By far the easiest way to get the Riemann tensor is to rewrite the metric as the product of vierbeins and obtain the spin connection one-forms [8]. This is straightforward using the property (14) and the fact that the other degrees of freedom depend only on the parameter t. From the spin connection we get the Riemann two-form, and the appropriate contractions of its components will yield the curvature scalar.

function as before and has the form,

$$H_{IX} = -\frac{1}{2a}p_a^2 + \frac{1}{2a^3}(p_+^2 + p_-^2) + \frac{1}{2}\lambda a^3 - \frac{a^3}{12}{}^3R_{IX} \tag{18}$$

upon identification of the momenta. This should be compared with (9). As you can see for yourself, the minisuperspace of the Bianchi-IX universe has only three dynamical degrees of freedom. If we further imposed isotropy we would have only one, and from the above the Hamiltonian for this FRW universe would be,

$$H_{FRW} = -\frac{1}{2a}p_a^2 + \frac{1}{2}\lambda a^3 - \frac{1}{2}a \tag{19}$$

4 Minisuperspace

It should be clear by now that minisuperspace has the advantage of reducing the potentially infinite number of degrees of freedom down to a manageable finite number. Imposing homogeneity on the spacelike slices removes the need for a degree of freedom at each point $\mathbf{x}$ on the slice. For the sake of clarity and ease of calculation let us assume homogeneity. The metric can therefore be expressed in terms of a few parameters q^α say,

$$ds^2 = -N^2\,\mathrm{d}t \otimes \mathrm{d}t + h_{ij}(q^\alpha)\,\mathrm{d}x^i \otimes \mathrm{d}x^j \tag{20}$$

The gravitational part of the action can then be expressed in terms of the q^α, their first time derivatives (courtesy of the boundary term $S_{\partial\mathcal{M}}$) and the lapse function as $S_G(q^\alpha, \dot{q}^\alpha; N)$. The momenta conjugate to these parameters are the usual $p_\alpha = \delta S/\delta\dot{q}^\alpha$ replacing the infinity of $\pi^{ij}(\mathbf{x})$. We can get the gravitational part of the Hamiltonian by variation of the action with respect to the lapse function,

$$H_G = \frac{1}{2\mu}f^{\alpha\beta}p_\alpha p_\beta + \mu U(q^\alpha) \tag{21}$$

where $\mu \sim M_P^2$. This is in a much simpler form than the full superspace expression (9). Just as it is sometimes convenient in quantum field theory to keep factors of $\hbar$ explicit rather than use natural units $\hbar = 1$ (in loop expansions for example) we have decided to keep factors of the Planck scale explicit here. We do so because we will later want to examine the WDW equation order by order in M_P^2. An indication of how to convert the q^α to naturalized QGD units can be seen by comparing (13) with (20) above.

The $f^{\alpha\beta}$ is the metric on the minisuperspace and is the analogue of the $G_{ij,kl}$. It is a function of the q^α, and has a signature of $(-++\ldots)$, the minus sign associated with the conformal scale factor $\sqrt{h}$. For example, in the Bianchi-IX model of the last section we have $f^{aa} = -1/a$ and $f^{++} = f^{--} = 1/a^3$. The QGD potential $U(q^\alpha)$ is just the intrinsic curvature and cosmological constant parts of (9) expressed in terms of the q^α. There are two problems we encounter when attempting to turn (21) into an operator. The first is that there will be an operator ordering ambiguity due to the q^α dependence of the metric when we replace the p_α by derivatives. Each possible operator ordering corresponds to a different quantum theory. The second is that there are many possible choices of parametrizing the three-geometry and of choosing the variables q^α; we could equally well have taken a^2 or a^3 as one of the variables and $a^2e^{\beta_\pm}$ as the other two in (13), rather than (a, β_+, β_-). Both these problems may be resolved at once [9, 10] if we demand that the Hamiltonian operator is the same no matter how we parametrize the h_{ij}, and thus all quantum theories are equivalent. This is done by demanding a general coordinate invariance on the minisuperspace. In this case we must make the substitution $f^{\alpha\beta}p_\alpha p_\beta \to -\nabla^2$, the Laplace operator on the minisuperspace,

$$\nabla^2 = \frac{1}{\sqrt{-f}}\frac{\partial}{\partial q^\alpha}\left(\sqrt{-f}f^{\alpha\beta}\frac{\partial}{\partial q^\beta}\right)$$

However, for now let us just treat ∇^2 as just a general derivative operator on the minisuperspace. In practice there is usually one particular operator ordering that makes the solution of the WDW straightforward as we shall see later.

The complete Hamiltonian is where we add on that for the matter, $H = H_G + H_\phi$, and this vanishes by the constraint from the lapse function. Thus the complete WDW equation is of the form,

$$\left(-\frac{1}{2\mu}\nabla^2 + \mu U(q) + \mathcal{H}_\phi\right)\Psi(q,\phi) = 0 \tag{22}$$

The assumption of homogeneity that makes the gravitational degrees of freedom finite in number has the same effect on the matter fields. These reduce to a few degrees of freedom ϕ, independent of position on the spacelike slice, and their conjugate momenta p_ϕ which replace the infinity of $\pi_\phi(\mathbf{x})$. The matter Hamiltonian is not only a function of the fields ϕ and their momenta p_ϕ but also of the metrical parameters, $H_\phi(p_\phi, \phi; q^\alpha)$. A glance at (9) and the example we gave of the scalar field should convince you of this[3]. The energy of the

[3] In general we include the matter degrees of freedom as coordinates q^α in the minisuperspace. The metric

matter is usually many orders of magnitude less than the Planck scale, but it is sizeable at the early stages of the universe. This is even so in the best cosmologies we have, based on inflation [11].

5 Solutions to the WDW equation

We can now examine the WDW equation (22) to find how classical behaviour emerges. There certainly ought to be some limit where we recover classical cosmology and quantum field theory for the matter on the resulting curved spacetime. This is remarkably straightforward [12] as we shall now explain. The kinetic term for the gravitational coordinates is suppressed by a factor of $\mu \sim M_P^2$ relative to the kinetic terms for the matter. We should therefore expect the development of the wave function to be slower in the q^a compared to the ϕ if we are in the regime where the matter energy is small compared to the Planck scale. If the wave function is in an eigenstate of the matter Hamiltonian then the slow evolution of the q^a will not cause a transition to another eigenstate. This is the criteria for the adiabatic or Born-Oppenheimer approximation [13]. In this approximation we write the wave function as $\Psi(q,\phi) = \psi(q)\chi(\phi;q)$. The $\chi(\phi;q)$ is the quasi-stationary matter wave function, a linear superposition of eigenstates of the matter Hamiltonian, $\mathcal{H}_\phi\chi_k(\phi;q) = E_k(q)\chi_k(\phi;q)$. When finding these eigenvectors we treat the q^a as constants, in keeping with the assumption of adiabaticity, and for this reason the eigenvalues $E_k(q)$ will end up with a dependence on them. The wave function for the geometry is assumed to be of the form $\psi(q) = e^{i\mu S_0}e^{iS_1}$ where the S_0 is real, meaning that the order μ-behaviour of $\psi(q)$ is purely oscillatory. What we are doing in this case is just the semi-classical or WKB approximation. Thus we have that,

$$\Psi(q,\phi) \sim e^{i\mu S_0}e^{iS_1}\chi(\phi;q) \tag{23}$$

is our approximation to the wave function. We can insert this into (22) and equate terms order by order in the Planck scale.

Equating the order μ terms gives us the Hamilton-Jacobi equation [14] for the gravitational degrees of freedom,

$f^{\alpha\beta}$ can then be extended to include the extra indices and the ∇^2 altered accordingly. We shall keep them separate for the moment as we shall be expanding in powers of the Planck scale.

$$\frac{1}{2}(\nabla S_0(q))^2 + U(q) = 0 \tag{24}$$

In other words, S_0 is an extremum of the classical gravitational action S_G above, with the upper end point having the coordinates q. The lower end point of the action is irrelevant here as it only contributes a constant to S_0. It would correspond to a different initial condition for the cosmology. Equating the terms to next order yields two equations since S_1 is complex. The imaginary part is,

$$\nabla S_0 \cdot \nabla(\Im S_1) = \frac{1}{2}\nabla^2 S_0 \tag{25}$$

which is just another expression of the WKB prefactor $e^{-\Im S_1}$ when solved. Thus S_0 and the imaginary part of S_1 together give the WKB approximation to the gravitational wave function. The real part (really the hermitian part) of the $o(1)$ terms is,

$$i\nabla S_0 \cdot \nabla\chi(\phi;q) = [\mathcal{H}_\phi + \nabla S_0 \cdot \nabla(\Re S_1)]\chi(\phi;q) \tag{26}$$

which tells us how the matter wave function evolves under the adiabatic evolution of the gravitation. As we said above, S_0 is an extremum of the classical action for the gravity. The extremum is attained on a classical path through the minisuperspace. The tangent vector to this path is therefore ∇S_0. We can parametrize the distance along the path by T say, and thus the derivative of anything along the length of the path is $\frac{d}{dT} = \nabla S_0 \cdot \nabla$. As we might have already guessed, the parameter that orders the classical path is just time. The coordinates of the path are $q^a(T)$, and they satisfy the classical equations of motion[4].

To see how the matter wave function evolves along this path we can re-express (26) as,

$$i\frac{d}{dT}\chi(\phi;q) = \mathcal{H}_\phi\chi(\phi;q) + \frac{d}{dT}(\Re S_1)\chi(\phi;q) \tag{27}$$

which indeed resembles the Schrödinger equation. We might think that we could absorb the last term into the phase of the wave function by redefining $\chi \to \chi' = e^{i\Re S_1}\chi$. This is also suggested by (23), whether the phase is in the gravitational or matter wave function being a question of choice. However, this is not so. The real part of S_1 is a non-integrable Berry

[4] As a very simple example we take the gravitational Hamiltonian of (19). The Laplacian can be written $\nabla^2 = -\frac{1}{a}\left(\frac{\partial}{\partial a}\right)^2$ in which case the Hamilton-Jacobi equation is $\frac{dS_0}{da} = -\sqrt{\lambda a^4 - a^2}$. Therefore the derivative along the classical path is $\frac{d}{dT} = \sqrt{\lambda a^2 - 1}\frac{d}{da}$. The value of the scale factor along the path is then $a(T) = \lambda^{-1/2}\cosh(\sqrt{\lambda}T)$ alias deSitter spacetime. You can check that S_0 has the correct sign from (16) with $\beta_\pm = 0$. If S_0 is the extremal action then ψ is the wave function for an *expanding* universe.

phase [15] which arises naturally from the use of the adiabatic approximation, although the only situation where it would manifest itself here would be for an oscillating universe. If the classical cosmology does not oscillate then we can redefine it away after all. Thus the matter evolves according to the familiar quantum field theory on the curved spacetime of the classical cosmology [16].

The above discussion has, of course, been restricted to where we are allowed to use the adiabatic and the WKB approximations. Naturally, we have recovered what we originally put in before developing the WDW equation (except, perhaps, for the Berry phase). However, this is only one solution to the WDW equation and a semi-classical one at that. There are many solutions to the WDW equation, and to select only one we have to impose boundary conditions on the wave function. This would then represent the true wave function of the universe. We are then forced to confront the question we have been avoiding up to now: How *do* we interpret the wave function of the universe? Whatever interpretation we adopt will differ from those we are familiar with in conventional quantum mechanics (where contention still exists, I might add) since 'time' is not inherent to the WDW equation, but rather arises out of the *solutions* when we happen to be in the semi-classical regime. Moreover, the standard 'observer' of quantum mechanics cannot exist without time to distinguish cause and effect (or preparation and measurement). This is just as well for quantum cosmology since there is no observer or measuring apparatus external to the universe, including Spinoza's god [17]. Moreover, reduction of the wave function of the universe by an internal measurement is patently nonsense.

What is known about solutions to the WDW equation is that in the WKB approximation the wave functions exhibit 'correlations' and 'decoherence' [18]. The former means that the wave function peaks on paths in minisuperspace corresponding to classical behaviour. The latter means that different classical paths do not interfere with each other. Once on a classical path we know from above that a time variable can be constructed. However, away from the classical regions, where the wave functions are not peaked, we cannot expect to understand what the wave function represents any more than we can understand what a tunnelling wave function under a barrier is. With the solution to the WDW equation fixed by some boundary condition there will be regions in the minisuperspace where it can be treated semi-classically. Each of these regions corresponds to a classical cosmology. This leads to one of the most

important applications of quantum cosmology, in questions concerning the initial state of the universe. The problem in classical cosmology, inflation included [11], is that we have to put in the initial conditions for it by hand at some stage after the Big Bang. The singularity prevents us from determining the initial conditions for the cosmology at the Big Bang itself. A different set of initial conditions will yield a different cosmology. With a given wave function it ought to be possible, in principle, to determine the most likely initial conditions for inflation [19] *i.e.* of all the possible classical cosmologies which is the one preferred. This has had some success, and for a review I refer you to Reference [20].

There is an interpretation of quantum mechanics that seems apt for quantum cosmology. This is the 'relative state' interpretation [21]. It allows for a wave function of the whole universe which does not collapse, and in this sense is deterministic. 'Observers' are internal subsystems of the universe, and as far as they are concerned it is quantum mechanics as usual. Their concept of time could be accommodated in this picture by having certain large subsystems act as clocks (although this has not been explicitly demonstrated in general). As we shall see, different applications of quantum comology usually provide their own interpretation. Also, different applications often require (or sometimes enforce) their own boundary conditions on the wave function and vice-versa. Recall that quantum mechanical scattering has different conditions at infinity compared to bound state problems. Two possible choices of boundary conditions in the literature that have received a great deal of attention give the 'Hartle-Hawking' wave function [2, 22] and the 'Vilenkin' wave function [23]. The boundary conditions are enforced at small and large three-geometries respectively, and they yield different predictions. Initial conditions for inflation, for example, are more readily derived from the latter [20].

What I am trying to get across is that there is nothing conceptually new here. Quantum cosmology employs the usual principles of quantization for its construction. Any 'understanding' is therefore limited by our understanding of quantization itself, and in this we still have a long way to go [24]. Quantum mechanics *per se* does not have much to say about its interpretation or boundary conditions. In the meantime we should be content that the rules of quantum mechanics, applied to a particular system, yield results agreeing with our observations of that system. Quantum cosmology should be evaluated by the same criteria. Given a cosmological system we apply the rules and calculate the consequences. There is no

unique interpretation or unique boundary condition valid for all situations. At this stage in its development QGD ought to be judged by its predictions and not its philosophy.

6 Path Integrals

The path integral (or sum-over-histories) approach to quantum mechanics complements the operator (Hamiltonian) approach [26]. As we all know, the former approach is more advantageous when quantizing gauge theories. As an alternative to the WDW equation we can construct a path integral for quantum cosmology. As there is a reparametrization invariance under different choices of the lapse and shift functions the kernel for quantum cosmology is derived in a manner similar to gauge theories. We have to gauge fix, introduce ghosts, and extract the 'gauge volume' as usual. However, as we shall see, there are a couple of interesting new points that arise. In the minisuperspace formulation we take the kernel to be,

$$K(q_f^a; q_i^a) = \int \mathcal{D}q^a\, \mathcal{D}p_a\, \mathcal{D}N \exp\left[i \int_i^f (p_a dq^a - N H dt)\right] \qquad (28)$$

where we have included the matter degrees of freedom as coordinates of the minisuperspace. The action has been written in its canonical form, and the Hamiltonian is of the form discussed in Section 4. We are integrating over all paths between given initial and final configurations of three-geometry and matter fields. The integration of the momenta is free at the end points. Also we integrate over all configurations of the lapse function. This is depicted in Figure 1.

As you can see from (28) we could consider performing the lapse functional integral to get a delta function of H. However, this is the wrong approach as we have to avoid the over-counting of paths equivalent by reparametrization. We have to choose a gauge. The simplest gauge-fixing condition is $\dot{N} = 0$, called the 'proper time' gauge for obvious reasons. How to proceed is described in Reference [10]. As expected we have to introduce ghost terms in the action. It can be shown that we can actually integrate out the ghost degrees of freedom. The result of this integration is a factor of $[t_f - t_i]$, where these are the values of the coordinate t at the end points. Moreover, the functional integral over the lapse $N(t)$ reduces to the integral over each choice of the constant $\int \mathcal{D}N(t) \to \int dN$. Therefore (28)

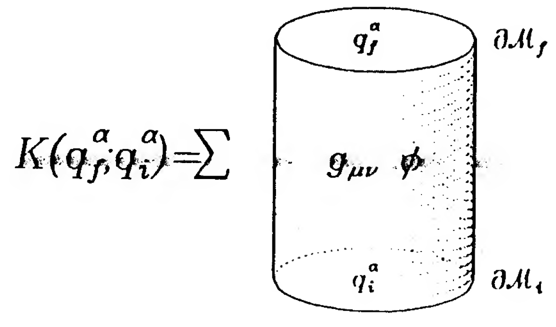

Figure 1: The kernel is the sum over all four-geometries and field configurations that interpolate between the specified values on the two boundaries.

reduces to,

$$K(q_f^a; q_i^a) = \int dN [t_f - t_i] \int \mathcal{D}q^a\, \mathcal{D}p_a \exp\left[i \int_i^f (p_a dq^a - N H dt)\right]$$

We can define a new variable $T = Nt$ (the proper separation between slices) in which case this further simplifies to the expression,

$$K(q_f^a; q_i^a) = \int d(T_f - T_i) \int \mathcal{D}q^a\, \mathcal{D}p_a \exp\left[i \int_{T_i}^{T_f} \left(p_a \frac{dq^a}{dT} - H\right) dT\right]$$

where $T_{f,i} = N t_{f,i}$ are the values at the boundaries. The remaining functional integral is formally equivalent to the familiar kernel of quantum mechanics, $\langle q_f^a, T_f | q_i^a, T_i \rangle$, where these states are in the Heisenberg representation [26]. As the Hamiltonian is independent of the T parameter we then know that $\langle q_f^a, T_f | q_i^a, T_i \rangle = \langle q_f^a, (T_f - T_i) | q_i^a, 0 \rangle$. Thus the quantum cosmology kernel is formally equivalent to,

$$K(q_f^a; q_i^a) = \int_{T_-}^{T_+} dT \, \langle q_f^a, T | q_i^a, 0 \rangle \qquad (29)$$

where for the moment we denote the upper and lower ranges of the T integral by $T_\pm$.

Our next question is whether this converges or not. First, we have not specified the range of integration of the T parameter. Secondly, we know that the Hamiltonian for QGD is unbounded from below due to the signature of the minisuperspace metric. This did not really bother us that much when we were interested in solving the WDW equation. Now that we have manipulated the kernel to the form (29) the bracket might not be well-defined

because of this. However, the form of (29) does seem quite reasonable and in keeping with our notions of time in quantum cosmology: we propogate from one configuration to the other, the interpolating path being parametrized by T, and then we sum over all the parametrizations.

To see if the bracket makes any sense let us insert a complete set of eigenstates of the WDW operator, $\mathcal{H}|n\rangle = E_n|n\rangle$, into it. What we find is that,

$$\langle q_f^\alpha, T|q_i^\alpha, 0\rangle = \sum_n \Phi_n(q_f^\alpha, T)\,\Phi_n^*(q_i^\alpha, 0)$$

where $\Phi_n(q^\alpha, T) = \langle q^\alpha, T|n\rangle$ are the wave functions of these eigenstates (those for which the eigenvalue $E_n = 0$ are solutions to the WDW equation). By definition of the Heisenberg representation we know that $\mathcal{H}\Phi_n(q^\alpha, T) = i\frac{\partial}{\partial T}\Phi_n(q^\alpha, T)$ and thus the bracket is equivalent to,

$$\langle q_f^\alpha, T|q_i^\alpha, 0\rangle = \sum_n e^{-iE_nT}\,\Phi_n(q_f^\alpha, 0)\,\Phi_n^*(q_i^\alpha, 0) \tag{30}$$

Ordinarily in field theory [26] we take $T \rightarrow \infty e^{-i\delta}$ so that the only state to survive the sum in (30) is the one of lowest energy, yielding the ground state to ground state amplitude in the Schrödinger representation. For quantum cosmology this is not appropriate as we have to integrate over the T parameter and, more importantly, there is no physical state of lowest energy. Our task is to find integration contours such that (29) converges. Just as in usual field theory where different choices for the contour in the complex momentum plane give Schwinger, Hadamard, Wightman, Feynman, advanced and retarded Green functions, so too will different contours in the complex T plane correspond to kernels with distinctive properties.

In order to understand the physical meaning of some of the contours it is quite useful to make a distinction between the matter and geometrical degrees of freedom as we did in Section 4. The minisuperspace coordinates are written $q^\alpha = (\tilde{q}^\alpha, \phi)$ and the Hamiltonian splits up into $H = H_G(\tilde{p}_\alpha, \tilde{q}^\alpha) + H_\phi(p_\phi, \phi; \tilde{q}^\alpha)$. The bracket in (29) is then equivalent to,

$$\langle \tilde{q}_f^\alpha, \phi_f, T|\tilde{q}_i^\alpha, \phi_i, 0\rangle = \int \mathcal{D}\tilde{q}^\alpha\,\mathcal{D}\tilde{p}_\alpha\, e^{iS_G} \int \mathcal{D}\phi\,\mathcal{D}p_\phi\, \exp\left[i\int_0^T \left(p_\phi \frac{d\phi}{dT'} - H_\phi\right) dT'\right]$$

where the action for the gravitational degrees of freedom is $S_G = \int_0^T \left(\tilde{p}_\alpha \frac{d\tilde{q}^\alpha}{dT'} - H_G\right) dT'$. The internal path integral over the matter is the kernel for the propagation of an initial field configuration, ϕ_i at $T' = 0$, to a final one, ϕ_f at $T' = T$, on the background metrical history

$[\tilde{q}^\alpha]$. Therefore we have,

$$\langle \tilde{q}_f^\alpha, \phi_f, T|\tilde{q}_i^\alpha, \phi_i, 0\rangle = \int \mathcal{D}\tilde{q}^\alpha\,\mathcal{D}\tilde{p}_\alpha\, e^{iS_G}\, G(\phi_f, T; \phi_i, 0; [\tilde{q}^\alpha]) \tag{31}$$

where $G(\phi_f, T; \phi_i, 0; [\tilde{q}^\alpha]) = (\phi_f, T; [\tilde{q}^\alpha]|\phi_i, 0; [\tilde{q}^\alpha])$, these being the Heisenberg basis for the development of the matter fields in the background spacetime. Equation (31) is seen to be the weighted sum of matter kernels over all interpolating geometries. The parameter T in (29) is therefore understood as the 'proper time' experienced by the matter on a particular history. This can be further clarified by acting on (31) with the matter Hamiltonian for the final configuration $\mathcal{H}_\phi^f = H_\phi(-i\frac{\partial}{\partial\phi_f}, \phi_f; \tilde{q}_f^\alpha)$ to have it act solely on the matter kernel,

$$\mathcal{H}_\phi^f\langle \tilde{q}_f^\alpha, \phi_f, T|\tilde{q}_i^\alpha, \phi_i, 0\rangle = \int \mathcal{D}\tilde{q}^\alpha\,\mathcal{D}\tilde{p}_\alpha\, e^{iS_G}\left(i\frac{\partial}{\partial T}\right) G(\phi_f, T; \phi_i, 0; [\tilde{q}^\alpha])$$

We have been able to pull $\mathcal{H}_\phi^f$ through the encompassing path integral since $\int \mathcal{D}\tilde{q}^\alpha$ does not integrate over the end point coordinate $\tilde{q}_f^\alpha$.

This identification of T with the proper time can be made clearer by considering the classical paths. For this we return to our arguments of Section 5 concerning the adiabatic approximation. Denote the eigenstates of the matter Hamiltonian by $\mathcal{H}_\phi|k;\tilde{q}^\alpha) = E_k(\tilde{q}^\alpha)|k;\tilde{q}^\alpha)$, $\tilde{q}^\alpha$ being regarded as externl parameters of the Hamiltonian. The wave functions for these states are $\chi_k(\phi, \tau; \tilde{q}^\alpha) = (\phi, \tau; [\tilde{q}^\alpha]|k; \tilde{q}^\alpha)$ and under the slowly developing geometry satisfy $i\frac{d}{d\tau}\chi_k(\phi, \tau; \tilde{q}^\alpha(\tau)) = E_k(\tilde{q}^\alpha(\tau))\,\chi_k(\phi, \tau; \tilde{q}^\alpha(\tau))$. The states are quasi-stationary, meaning that under the slow change of the geometry there are no transitions to other eigenstates. The time development of the wave functions is given by [15],

$$\chi_k(\phi, \tau; \tilde{q}^\alpha(\tau)) = \exp\left[-i\int_0^\tau E_k(\tilde{q}^\alpha(\tau'))d\tau'\right] e^{i\gamma_k(\tau)}\chi_k(\phi, 0; \tilde{q}^\alpha(\tau))$$

where the extra term (the Berry phase) is in principle non-integrable and is given by $\gamma_k(\tau) = i\int_0^\tau \left(\frac{d\tilde{q}^\alpha}{d\tau'}\right)(k;\tilde{q}|\nabla_\alpha|k;\tilde{q})d\tau'$. The state $\chi_k(\phi_i, 0; \tilde{q}_i^\alpha)$ propagates to the state $\chi_k(\phi_f, T; \tilde{q}_f^\alpha)$. The matter kernel of (31) can therefore be expressed as the sum,

$$G(\phi_f, T; \phi_i, 0; [\tilde{q}^\alpha]) \approx \sum_k \exp\left[-i\int_0^T E_k(\tilde{q}^\alpha(T'))dT'\right] e^{i\gamma_k(T)}\,\chi_k(\phi_f, 0; \tilde{q}_f^\alpha)\chi_k^*(\phi_i, 0; \tilde{q}_i^\alpha) \tag{32}$$

in the adiabatic approximation, to be compared with (30). Also in this approximation the gravitational action $S_G([\tilde{q}^\alpha])$ is implicitly of order the Planck scale. Thus the remaining path

integral is dominated by the classical path, for which S_G is a minimum. Denoting this path by $\tilde{q}_0^\alpha(T'')$ we find that the WKB approximation to (31) is,

$$\langle \tilde{q}_f^\alpha, \phi_f, T | \tilde{q}_i^\alpha, \phi_i, 0 \rangle \sim \Delta\, e^{iS_0}\, G(\phi_f, T; \phi_i, 0; [\tilde{q}_0^\alpha]) \tag{33}$$

where S_0 is the gravitational action for the classical path between the initial and final configurations of the geometry, $S_0(\tilde{q}_f^\alpha, T; \tilde{q}_i^\alpha, 0) = S_G([\tilde{q}_0^\alpha])$. The prefactor Δ is the determinant of the fluctuations around the classical path and can be written [26] as,

$$\Delta = \left[\det\left(\frac{i}{2\pi}\frac{\partial^2 S_0}{\partial \tilde{q}_f^\alpha \partial \tilde{q}_i^\beta}\right)\right]^{1/2}$$

Along the classical path we also have $\frac{d}{dT'} = \frac{d\tilde{q}_0^\alpha}{dT}(T'')\nabla_\alpha = \nabla S_0(T'') \cdot \nabla$ where $S_0(T'')$ is the action on the classical path integrated up to the intermediate time T''. This makes explicit the identification of the time on the classical path with the integration parameter T of (29).

Let us now return to the contours of T integration. As we said, the bracket of (29) is divergent as $T \to \infty e^{-i\delta}$ due to the absence of any lowest energy state in the theory. An ingenious suggestion was made by Linde [23] that perhaps we could consider taking $T \to \infty e^{+i\delta}$ to pick out the state of *highest* energy. However, the circumstances under which such a state can exist are seldom met in practice and so this approach has limited use. There are two choices that are immediately apparent. The first is to take $T_\pm = \pm\infty$ in (29) so that we only pick out the states with $E_n = 0$ [27]. This kernel is then such that the WDW operator acting on either argument yields zero. The second is to take the half-infinite range $T_- = 0, T_+ = \infty$ and to add [25, 28] a Feynman $-i\epsilon$ to the Hamiltonian. This introduces a factor of $\exp(-\epsilon T)$ into (29) which makes the integral converge at the upper limit. It is then a simple task to show that the WDW operator acting on this kernel gives[5] $(\mathcal{H}^f - i\epsilon)K_+(q_f^\alpha; q_i^\alpha) = -i\delta(q_f^\alpha; q_i^\alpha)$, and this has been likened to a retarded propagator by Teitelboim [28] as we are integrating over paths for which the final configuration lies to the *future* of the initial configuration.

The original suggestion of Hartle and Hawking [22] was to Wick rotate $T \to -i\tau$ to define a path integral over Euclidean metrics. Their motivation for this was that Euclidean

[5] The form of the delta function depends on the metric of the minisuperspace in the usual way, $\delta(q_f^\alpha; q_i^\alpha) = (-f)^{-1/2}\prod_\alpha \delta(q_f^\alpha - q_i^\alpha)$ so that upon integration by the natural volume element $\sqrt{-f}\prod dq^\alpha$ it yields unity

metrics can be closed and compact (free of singularities), and consequently a boundary condition could be imposed at vanishing three-geometry. This is the 'no boundary' proposal. The propagator is then a wave function $\Psi_{HH}(q_f^\alpha) = K_E\left(q_f^\alpha; (\det h)_i = 0\right)$. However, the immediate problem is that the Euclidean action of gravity is unbounded from below. Once again we have to define an integration contour in complex τ-space to define the path integral.

The fact that generally we have to take the integral of (29) to be a contour in complex T-space is hardly a surprise. The same thing happens in usual quantum mechanics when we have to accommodate barrier penetration in a path integral setting [29]. In our case there are regions of minisuperspace that do not admit classical cosmologies. If a path passes through them we are required to go off the real axis and include complex four-geometries. This question of contours is related to the comments at the end of Section 5 concerning boundary conditions. Depending on which problem we are investigating we choose our boundary conditions appropriately. Having done so we are still left with a choice of T contour to define the propagator. There are many contours that have been examined in the literature [30], each with their own distinctive properties. We are in need of a deeper physical principle to tell us which of the QGD propagators is in a sense more fundamental than the others. Nevertheless, convergent path integrals can be defined.

7 The Cosmological Constant

But it's strange ... This nothing that weighs so much.

Aristophanes

A long-standing mystery in particle physics and gravitation is why Einstein's cosmological constant is so many orders of magnitude smaller than it ought to be [31]. The particle physics vacuum inevitably has some energy locked up in it. It is this residual energy density, measured at energy scales far below those of the underlying particle dynamics, that is the cosmological constant. The upper limit on this energy density is on the order of $10^{-47}\mathrm{GeV}^4$ from our present knowledge of the large scale structure of the universe. This is a highly unnatural scale from the standpoint of particle physics. To arrange for a cancellation by the necessary amount requires a fine tuning of the particle physics, and all its vacuum fluctuations, to an incredible degree of accuracy. Even if it is possible to succeed in this, all that has happened is that we have exchanged one problem for a host of others. We are in serious need of some mechanism for neutralizing the vacuum energy of the universe.

If quantum cosmology truly is the key to understanding the properties of our universe then it certainly ought to explain why the cosmological constant is so small. There are several ways to approach the problem. The most notable in the past few years has been the wormhole program which we shall be exploring later. However, I must emphasize that there are other, less radical, mechanisms available. In the next section I shall discuss the method of 'third quantization'. The notion of using quantum cosmology to solve the cosmological constant problem goes back to an observation of Baum [32] and Hawking [33]. In Euclidean quantum gravity the path integral has a stationary value of $\exp(3\pi/G_N\Lambda)$. If the cosmological constant Λ can somehow be made a dynamical variable then the Euclidean path integral would be dominated by geometries for which $\Lambda \to 0_+$. This (BH) factor crops up quite often in quantum cosmology in a variety of contexts. As we discussed in Section 5, the WDW equation typically has classically allowed and forbidden regions. It is possible to calculate the tunnelling probability through a classically forbidden region and, depending on the boundary conditions, it can be the BH factor [22] or its reciprocal [23]. However, in the above cases Λ is a fixed parameter of the theory and therefore the BH mechanism is not directly applicable.

To understand how the cosmological constant can first be made a 'variable' and can then have its value driven to zero it is useful to examine a simple model. It illustrates all the steps that enter into the various mechanisms. The model was originally used by Hawking [33] and has since undergone some development [34, 35]. It makes use of the special properties of a four-form (fourth rank antisymmetric tensor) field in four dimensions. It has been known for some time that the presence of a four-form gives rise to a variable cosmological constant [36, 37]. Its physics is straightforward and, most important of all, we can calculate with it consistently.

A four-form's action is similar to the Maxwell action for electromagnetism, but with a couple of extra indices. In Euclidean space it is,

$$S_F = \frac{1}{2}\int_{\mathcal{M}} \mathbf{F}\wedge *\mathbf{F} = \frac{1}{2}\frac{1}{4!}\int_{\mathcal{M}} F_{\mu\nu\lambda\rho}F^{\mu\nu\lambda\rho}\sqrt{g}\,d^4x \qquad (34)$$

Such a field arises naturally in theories with supergravity [36], although for our purposes it is simply an introduced field with no particular theory in mind. Let us denote the vacuum energy that we wish to cancel by $\Lambda_0/8\pi G_N$ i.e. represent it by a 'bare' cosmological constant. The total Euclidean action for the four-form coupled to gravity is then,

$$S_E = -\frac{1}{16\pi G_N}\int_{\mathcal{M}}(R-2\Lambda_0)\sqrt{g}d^4x + S_F \qquad (35)$$

The Euclideanized version of a closed FRW metric is where we take (12) with $k = 1$ and flip the sign of the dt^2 term. For a positive cosmological constant Λ the solution to the Euclidean Einstein equations is that the spacetime is a four-sphere of radius $\sqrt{3/\Lambda}$. This is a closed space which is why the boundary contributions can be dropped in (35).

The key to studying the four-form on a closed space is the Hodge decomposition [8]. It allows us to uniquely decompose any form into an orthogonal sum of exact, co-exact and harmonic forms. Thus we can write $\mathbf{F} = \mathbf{dA} + \kappa\boldsymbol{\eta}$ where $\mathbf{A}$ is a globally defined three form and $\boldsymbol{\eta}$ is the (harmonic) volume form on the manifold. The co-exact contribution is absent in this case as there are no co-exact four-forms on a four-dimensional manifold. The coefficient κ is a constant on the manifold. The action (34) is therefore equivalent to,

$$S_F = \frac{1}{2}\int_{\mathcal{M}}\mathbf{dA}\wedge *\mathbf{dA} + \frac{1}{2}\kappa^2\int_{\mathcal{M}}\boldsymbol{\eta} \qquad (36)$$

We treat the components of $\mathbf{A} = \frac{1}{3!}A_{\mu\nu\lambda}\,\mathbf{d}x^\mu\wedge\mathbf{d}x^\nu\wedge\mathbf{d}x^\lambda$ as the only dynamical degrees of freedom of $\mathbf{F}$. The parameter κ is taken to arise from a choice of background value for the

four form (we need not treat it as a dynamical degree of freedom as it has no kinetic term). In components, $F_{\mu\nu\lambda\rho} = f_{\mu\nu\lambda\rho} + \kappa\epsilon_{\mu\nu\lambda\rho}$ where the latter is the permutation tensor (not the density) and where f is the curl of A, $f_{\mu\nu\lambda\rho} = \partial_\mu A_{\nu\lambda\rho} - \partial_\nu A_{\lambda\rho\mu} + \partial_\lambda A_{\rho\mu\nu} - \partial_\rho A_{\mu\nu\lambda}$. The stress tensor is obtained as usual from the variation of the action with respect to the metric,

$$T_{\mu\nu} = \frac{1}{4!}\left(4 f_{\mu\lambda\rho\sigma} f_\nu{}^{\lambda\rho\sigma} - \frac{1}{2} g_{\mu\nu} f_{\lambda\rho\sigma\tau} f^{\lambda\rho\sigma\tau}\right) - \frac{1}{2}\kappa^2 g_{\mu\nu} \tag{37}$$

The term in brackets is essentially the Maxwell stress tensor, but with an extra couple of indices.

Varying the action (35) with respect to the metric will give the usual Einstein equations, $G_{\mu\nu} = 8\pi G_N T_{\mu\nu} - \Lambda_0 g_{\mu\nu}$ and variation of the action with respect to $A_{\mu\nu\lambda}$ gives the equation of motion for the three-form $\mathbf{d} * \mathbf{dA} = 0$. This would imply that $\mathbf{dA} = \kappa'\boldsymbol{\eta}$ with κ' a constant as before, only that the Hodge decomposition is unique, **A** is not harmonic, and thus we must have $\kappa' = 0$. The background value of the stress tensor is therefore $T_{\mu\nu} = -\frac{1}{2}\kappa^2 g_{\mu\nu}$ as $\mathbf{f} = 0$. Therefore the Einstein equations are simply $G_{\mu\nu} = -\Lambda(\kappa) g_{\mu\nu}$ where the effective cosmological constant is $\Lambda(\kappa) = \Lambda_0 + 4\pi G_N \kappa^2$, and upon contraction we find that the curvature scalar is $R = 4\Lambda(\kappa)$. The action (35) for this classical solution is then $\tilde{S}_E = -\Lambda(\kappa)V/8\pi G_N$ where $V = \int \sqrt{g} d^4x$ is the volume of the four sphere. The latter quantity is $24\pi^2/\Lambda(\kappa)^2$ and so we find that the stationary value of the Euclidean action for this system is just $\tilde{S}_E = -3\pi/G_N\Lambda(\kappa)$. Note that this value is negative.

The parameter κ characterizes the background configuration of the four-form and is not determined by anything in the theory; it is external to the theory. The cosmological constant is dependent on which background we are in. If we are willing to accept that the semi-classical approximation to the Euclidean path integral is $Z_E \sim \exp(-\tilde{S}_E)$, then in the κ-vacuum it is $Z_E(\kappa) \sim \exp(3\pi/G_N\Lambda(\kappa))$. We can use the path integral to calculate expectation values, which will also depend on the vacuum we are in. We have achieved our first goal of allowing different values for the cosmological constant, but as yet its value is fixed external to the theory. The second element of the BH mechanism appears to be an appeal to a *deus ex machina*. It is that there exists another theory, bigger than the one we were considering, that has the above theories as distinct subsectors, with a larger generating functional something like $Z_E \sim \int \mathcal{D}\kappa\, Z_E(\kappa)$. Then, due to the BH factor, this is dominated by the theory for which $\Lambda(\kappa) \to 0_+$. If this sounds a little vague it is because the

mechanism works slightly differently for the two cases we shall be considering here, those of third quantization and wormholes.

8 Third Quantization

As we saw in an earlier section, the WDW equation has similarities to the more familiar Klein-Gordon (KG) equation. Not surprisingly there is also a difficulty with normalization. Ordinarily this is remedied by promoting the KG wave function to the status of a field operator. We do the same to the wave function of the universe [38, 39, 40]. This field operator satisfies the WDW equation and we have now second quantized the gravity. As the matter was previously second quantized this explains why the procedure has been dubbed 'third' quantization. Because we are now in a different Hilbert space we have our *deus ex machina*. We shall show that what the quanta of the field operator correspond to are different particle physics vacua, and that the dominant one has vanishing cosmological constant. The model I'll describe here is one recently devised by Lars Jensen and myself [41] and it pulls together most of the results of the previous sections. This is a Lorentzian model, but we shall find that we recover the BH factor of the above Euclidean analysis by quite a different route.

We take our spacetime to be closed FRW, in which case the metric is (13) with the $\beta_{ij} = 0$. The gravitational Hamiltonian is then (19). We have some generic particle physics model with a vacuum state $|\Omega_0\rangle$ of lowest energy. This has a vacuum energy density ρ_0 say, which we write as a bare cosmological constant Λ_0 as before. When scaled as in Section 3 it is λ_0 in Planck units. In addition to the generic particle physics model we have a four-form which will allow for different values of the effective cosmological constant. The four-form is written as the curl of a three-form potential $\mathbf{F} = \mathbf{dA}$. As the spacelike slices are homogeneous we can expand the potential in its basis,

$$\mathbf{A} = \chi(t)\,\boldsymbol{\sigma}^1 \wedge \boldsymbol{\sigma}^2 \wedge \boldsymbol{\sigma}^3 + \frac{1}{2}\zeta_{ij}(t)\,\mathbf{d}t \wedge \boldsymbol{\sigma}^i \wedge \boldsymbol{\sigma}^j$$

When we take the curl we find that **F** is independent of ζ_{ij}. There is only one degree of freedom for a four-form in general, and in our particular case it is $\mathbf{F} = \dot{\chi}\,\mathbf{d}t \wedge \boldsymbol{\sigma}^1 \wedge \boldsymbol{\sigma}^2 \wedge \boldsymbol{\sigma}^3$.

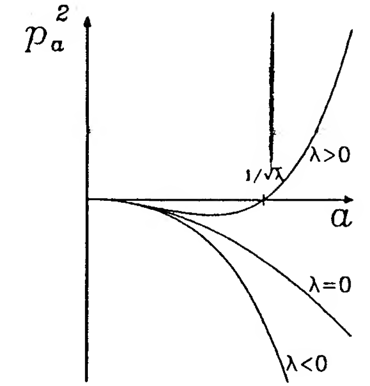

Figure 2: The behaviour of the momentum conjugate to the scale factor for different effective cosmological constants. Classical Lorentzian solutions (deSitter space) only exist for $\lambda > 0$.

The action for the four-form in a Lorentzian spacetime is,

$$S_F = -\frac{1}{2}\int_{\mathcal{M}} \mathbf{F}\wedge *\mathbf{F} = -\frac{1}{2}\frac{1}{4!}\int_{\mathcal{M}} F_{\mu\nu\lambda\rho}F^{\mu\nu\lambda\rho}\sqrt{-g}\,d^4x \tag{38}$$

You can also see that the action (38) results in only one degree of freedom for $\mathbf{F}$ in much the same way that you identify those of the Maxwell field in QED; essentially the ζ_{ij} may be gauged away. If we express the four-form degree of freedom in Planck units via the rescaling $\chi = (2\pi^2)^{-1/2}\left(\frac{2G_N}{3\pi}\right)\phi$ the action (38) is equivalent to $S_F = \frac{1}{2}\int dt\left(\frac{\dot{\phi}^2}{Na^3}\right)$. From this we can find the matter Hamiltonian $H_\phi = \frac{1}{2}a^3p_\phi^2$. Adding this to (19) the total Hamiltonian is then,

$$H = \frac{1}{2}\left(-\frac{p_a^2}{a} + a^3p_\phi^2 - a + \lambda_0 a^3\right) \tag{39}$$

which is our starting point. The classical behaviour is that p_ϕ is conserved (there is no potential for ϕ) and that the cosmology is deSitter with an effective cosmological constant $\lambda = \lambda_0 + p_\phi^2$. Thus the four-form is $\mathbf{F} \sim p_\phi\eta$ classically. Comparing this to the the Euclidean arguments of the last section we see that the momentum p_ϕ is analogous to the undetermined constant κ. The behaviour of p_a^2 is shown in Figure 2.

The value of p_ϕ is arbitrary and characterizes the vacuum configuration $|p_\phi\rangle$ of the four-form. There is an infinity of possible configurations. The combined matter and four-form vacuum is $|\Omega_0; p_\phi\rangle$, and in this vacuum the effective cosmological constant is $\lambda = \lambda_0 + p_\phi^2$. As λ_0 is usually negative (particle physics vacua usually have a negative energy density from the effects of symmetry breaking) we observe that the special states $|\Omega_0; \pm\sqrt{-\lambda_0}\rangle$ have vanishing

effective cosmological constant. However, we have to be in exactly the right states for this to happen. Third quantization is going to select these states for us.

By the procedures discussed in previous sections we can construct the WDW operator from (39). This can be written as,

$$\mathcal{H} = \frac{1}{2a}\left[a^{-n}\frac{\partial}{\partial a}\left(a^n\frac{\partial}{\partial a}\right) - a^4\frac{\partial^2}{\partial\phi^2} + (\lambda_0 a^4 - a^2)\right] \tag{40}$$

where the factor ordering ambiguity is here represented by the parameter n. The WDW equation is the usual $\mathcal{H}\Psi(a,\phi) = 0$. The scale parameter a and the four-form ϕ are analogous to time and a spatial coordinate in the KG equation respectively. The last term in (40) is analogous to the mass or potential of the KG equation. However, its dependence on the scale factor (the pseudo-time coordinate) means that it resembles more the KG equation on a curved background space [16]. Now it follows from (40) and the WDW equation that there is a conserved (minisuperspace) current whose components are,

$$J^a = ia^n\Psi^* \overleftrightarrow{\partial}_a \Psi$$
$$J^\phi = -ia^{n+4}\Psi^* \overleftrightarrow{\partial}_\phi \Psi$$

and which satisfy $\partial_a J^a + \partial_\phi J^\phi = 0$. We can use this to define an inner product of any two wave functions which are solutions to the WDW equation according to,

$$(\Psi_1|\Psi_2) \equiv ia^n\int_{-\infty}^{\infty}\left(\Psi_1^* \overleftrightarrow{\partial}_a \Psi_2\right)d\phi \tag{41}$$

Both wave functions are evaluated at the same value of the scale. Under this inner product the operator $p_\phi = -i\frac{\partial}{\partial\phi}$ is hermitean, as is the 'frequency' $\omega = ia^n\frac{\partial}{\partial a}$. We use these to define our states.

We can now construct a complete set of orthonormal solutions to the WDW equation. With a judicious choice of $n = -1$ for the factor ordering parameter we would find that these are of the form,

$$u_k(a,\phi) = \mathcal{N}_k e^{ik\phi}\left[\mathrm{Ai}(z) + \beta_k\mathrm{Bi}(z)\right] \tag{42}$$

where $\mathrm{Ai}(z)$ and $\mathrm{Bi}(z)$ are Airy functions [42]. They are linearly independent solutions to the differential equation $w''(z) - zw(z) = 0$. The k is the eigenvalue of p_ϕ and labels the vacuum state of the four-form. For the operator (40) you can check that the argument of the Airy functions is for each k state is,

$$z = \frac{1 - \lambda_k a^2}{(4\lambda_k^2)^{1/3}}$$

where $\lambda_k = \lambda_0 + k^2$ is the effective cosmological constant for that state. The boundary conditions we shall impose on the wave functions will determine the parameter β_k. The Wronskian of the Airy functions is $\mathrm{Ai}(z)\mathrm{Bi}'(z) - \mathrm{Ai}'(z)\mathrm{Bi}(z) = 1/\pi$ which fixes the normalization from (41). If the u_k are delta function normalizable then the factor $\mathcal{N}_k$ is determined by,

$$\frac{8|\mathcal{N}_k|^2}{(4\lambda_k^2)^{1/3}}\left[\lambda_k \Im\beta_k\right] = 1$$

and immediately we see that $\Im\beta_k$, the imaginary part of β_k, must have the same sign as λ_k. As expected, our wave functions have to be complex otherwise they cannot be normalized. Under the inner product (41) this complete set of orthonormal solutions to the WDW equation satisfies,

$$\begin{aligned} (u_k|u_{k'}) &= \delta(k-k') \\ (u_k^*|u_{k'}) &= 0 \\ (u_k^*|u_{k'}^*) &= -\delta(k-k') \end{aligned} \tag{43}$$

as can be easily verified. Each mode u_k is a solution to the WDW equation and is the wave function for a universe with vacuum $|\Omega_0; k)$.

The next step of third quantization is to promote the wave function Ψ to the status of an operator and expand it in terms of the modes u_k,

$$\hat{\Psi}(a,\phi) = \int_{-\infty}^{\infty} dk \left[\hat{c}(k)u_k(a,\phi) + \hat{c}^\dagger(k)u_k^*(a,\phi)\right] \tag{44}$$

The $\hat{c}$ and $\hat{c}^\dagger$ are the annihilation and creation operators for each mode and are taken to satisfy the familiar canonical commutation relations, $\left[\hat{c}(k), \hat{c}^\dagger(k')\right] = \delta(k-k')$ etc. The equal scale commutation relation for the operator $\hat{\Psi}$ can then be found,

$$\left[ia^n\frac{\partial}{\partial a}\hat{\Psi}(a,\phi), \hat{\Psi}(a,\phi')\right] = \delta(\phi-\phi')$$

in a similar way to (43). This entire procedure is analogous to the quantization of the KG scalar field on a curved spacetime as noted above[6]. We have yet to determine the

[6] Equally well, instead of canonically quantizing we could construct a path integral for the third quantization [40]. For the model considered here the action for this would be,

$$S_3 = \frac{1}{2}\int a^n da d\phi \left[\left(\frac{\partial\Psi}{\partial a}\right)^2 - a^4\left(\frac{\partial\Psi}{\partial\phi}\right)^2 - (\lambda_0 a^4 - a^2)\Psi^2\right]$$

which yields the WDW equation upon variation with respect to Ψ.

parameters β_k. Recall that the frequency operator $\omega = ia^n\frac{\partial}{\partial a}$ is hermitean. You can show by the methods of Section 5 that in classical regimes, where the wave function has the WKB form (23), that a positive frequency mode corresponds to an expanding classical universe. The boundary condition we impose on the wave functions, even in the non-classical regions, is that the modes u_k have a positive frequency whereas their conjugates u_k^* have negative frequency. However, as the potential in (40) is scale dependent so will be the meaning of positive frequency. We therefore have many possible expansions (44) depending on at which scale we choose to fix the frequency. We shall consider the two natural choices of small scales ($a \gtrsim 1$) and large scales ($a \to \infty$), and impose the positive frequency conditions there. We call these the 'in' and 'out' modes in keeping with the nomenclature for the KG field [16].

The effective cosmological constant can have either sign depending on the magnitude of the wavenumber k. For a $\lambda_k \leq 0$ there are no classical cosmologies, but we can still construct positive frequency states. In this case it turns out that if the modes are positive frequency at small scales they are also so at large scales. Thus the 'in' and 'out' modes are identical for $\lambda_k \leq 0$. For the special case of $k = \pm k_0 = \pm|\lambda_0|^{1/2}$, when $\lambda_k = 0$, they are simply,

$$u^{\mathrm{in}}_{\pm k_0}(a,\phi) = u^{\mathrm{out}}_{\pm k_0}(a,\phi) = \frac{e^{\pm ik_0\phi}}{\sqrt{8\pi}}\left(e^{a^2/2} + ie^{-a^2/2}\right) \tag{45}$$

as can be seen by solving the WDW equation. The normalization is fixed by (43). With this choice of wave function the phase of (45) has a scale dependence of $\vartheta = -\arctan\left[\tanh\left(\frac{a^2}{2}\right)\right]$ leading to a positive frequency $\omega(a) = -\frac{1}{a}\frac{\partial\vartheta}{\partial a} = \mathrm{sech}(a^2)$. For $|k| < k_0$, where $\lambda_k < 0$, the modes have the form,

$$u^{\mathrm{in}}_k(a,\phi) = u^{\mathrm{out}}_k(a,\phi) = \frac{e^{ik\phi}}{(16|\lambda_k|)^{1/6}}\left[ie^{1/3|\lambda_k|}\mathrm{Ai}(z) + \frac{1}{2}e^{-1/3|\lambda_k|}\mathrm{Bi}(z)\right] \tag{46}$$

The frequency of these modes is also $\omega(a) = \mathrm{sech}(a^2)$, and from the asymptotic properties of the Airy functions for large positive argument [42] it can be seen that these converge to the modes (45) in the limit of $k \to \pm k_0$. For the case of a positive effective cosmological constant $\lambda_k > 0$ the 'in' and 'out' modes are distinct. They are respectively,

$$u^{\mathrm{in}}_k(a,\phi) = \frac{e^{ik\phi}}{(16\lambda_k)^{1/6}}\left[e^{1/3\lambda_k}\mathrm{Ai}(z) + \frac{i}{2}e^{-1/3\lambda_k}\mathrm{Bi}(z)\right] \tag{47}$$

$$u^{\mathrm{out}}_k(a,\phi) = \frac{e^{ik\phi}}{(128\lambda_k)^{1/6}}\left[\mathrm{Ai}(z) + i\mathrm{Bi}(z)\right] \tag{48}$$

The former is only positive frequency at small scales where it has $\omega(a) = \mathrm{sech}(a^2)$, and it also converges to the modes (45) in the limit $k \to \pm k_0$. The latter is only positive frequency at large scales, where it is the wave function for a deSitter universe with positive effective cosmological constant. The reason for the distinction between the two modes here is that there is a clear separation between the regions $a > 1/\sqrt{\lambda_k}$ and $a < 1/\sqrt{\lambda_k}$.

The 'in' and 'out' modes are two different sets of complete orthonormal solutions to the WDW equation, either of which may be used as the basis for a mode expansion of the wave operator. The 'in' modes taken together are (46),(45) and (47) in order of increasing absolute wavenumber, whereas the 'out' modes are (46), (45) and (48). The wave operator can be expanded in terms of either set,

$$\begin{aligned} \hat{\Psi}(a,\phi) &= \int dk \left[\hat{c}_{\mathrm{in}}(k) u_k^{\mathrm{in}}(a,\phi) + \hat{c}_{\mathrm{in}}^{\dagger}(k) u_k^{\mathrm{in}*}(a,\phi)\right] \\ &= \int dk \left[\hat{c}_{\mathrm{out}}(k) u_k^{\mathrm{out}}(a,\phi) + \hat{c}_{\mathrm{out}}^{\dagger}(k) u_k^{\mathrm{out}*}(a,\phi)\right] \end{aligned} \tag{49}$$

each set with their own annihilation and creation operators. Each set of modes has its own (third quantized) vacuum, which is annihilated by their respective operators,

$$\hat{c}_{\mathrm{in}}(k)\,|0,\mathrm{in}\rangle\!\rangle = 0\,, \quad \hat{c}_{\mathrm{out}}(k)\,|0,\mathrm{out}\rangle\!\rangle = 0 \tag{50}$$

for all k. The double ket signifies that these are third quantized Fock vacua, not to be confused with the second quantized particle physics Fock vacuum $|\Omega_0\rangle$. We now have the means to determine the dominant particle physics vacuum $|\Omega_0;k\rangle$. Given that the third quantized state of the theory is $|\Omega_3\rangle\!\rangle$ we can estimate the expectation value,

$$\mathcal{N}(k,k') = \langle\!\langle\Omega_3|\hat{c}_{\mathrm{out}}^{\dagger}(k)\hat{c}_{\mathrm{out}}(k')|\Omega_3\rangle\!\rangle \tag{51}$$

This is the distribution of k-vacua for a very large universe such as ours. A large universe has a scale factor $a \gg 1$ in Planck units, which is why we are interested in the number of 'out' modes. What we mean by a small universe is one having a scale factor $a \gtrsim 1$; for scale factors such as these, sufficiently far above the Planck length, the frequency $\omega(a) \ll 1$ and so the wave functions do not have fluctuations of the order of the Planck length itself. As an ansatz for the state vector we take $|\Omega_3\rangle\!\rangle = |0,\mathrm{in}\rangle\!\rangle$. This is a choice frequently encountered in the literature, sometimes without justification. Rubakov has pointed out [38] that such a state, or one closely related to it, seems to be crucial in determining the behaviour at large

scale factors. Here we can partly justify our ansatz by noting that our effective theory should break down at scale factors of order the Planck length. Since such small universes would be poorly defined we can require their absence in the third quantized state.

In order for us to calculate the expectation of 'out' operators in the 'in' vacuum we have to transform between the two bases of modes. The Bogolubov coefficients [43] are obtained by expanding the 'out' modes in terms of the complete set of 'in' modes,

$$u_k^{\mathrm{out}}(a,\phi) = \int_{-\infty}^{\infty} dq \left[\alpha(k,q) u_q^{\mathrm{in}}(a,\phi) + \beta(k,q) u_q^{\mathrm{in}*}(a,\phi)\right] \tag{52}$$

By the orthonormality conditions (43) we identify the coefficients as the inner products,

$$\alpha(k,q) = (u_q^{\mathrm{in}}|u_k^{\mathrm{out}})\,, \quad \beta(k,q) = -(u_q^{\mathrm{in}*}|u_k^{\mathrm{out}}) \tag{53}$$

It is a simple matter to calculate these for the expressions we have found above. By taking inner products of (49) with the various modes we can similarly express,

$$\hat{c}_{\mathrm{out}}(k) = \int dq \left[\alpha^*(k,q)\hat{c}_{\mathrm{in}}(q) - \beta^*(k,q)\hat{c}_{\mathrm{in}}^{\dagger}(q)\right]$$

and likewise for the conjugate. Inserting this expansion into (51) between the 'in' vacuum we would find that the number of states $\mathcal{N}(k,k')$ is solely determined by the coefficient $\beta(k,q)$,

$$\mathcal{N}(k,k') = \int dq\, \beta^*(k,q)\beta(k',q)\,.$$

The third quantized state may be devoid of modes at small scales, but it does contain modes at large scales. Upon calculating the β coefficients we find that,

$$\mathcal{N}(k,k') = \begin{cases} \frac{1}{2}\left(e^{1/3\lambda_k} - \frac{1}{2}e^{-1/3\lambda_k}\right)^2 \delta(k-k') & \text{if } \lambda_k > 0 \\ 0 & \text{if } \lambda_k \leq 0 \end{cases} \tag{54}$$

where the latter is just a statement of the two bases agreeing on their modes (45) and (46) i.e. $\beta(k,q) = 0$ for $|k| \leq k_0$. The first expression is strongly peaked as $\lambda_k \to 0_+$ and the density of k-vacua is then,

$$\rho(k) \sim e^{2/3\lambda_k}\Theta(\lambda_k) = \exp\left(\frac{3\pi}{G_N\Lambda(k)}\right)\Theta(\Lambda(k)) \tag{55}$$

precisely the Baum-Hawking factor. The density is sketched in Figure 3. Therefore the third quantized state of the universe is dominated by the quanta that correspond to a large

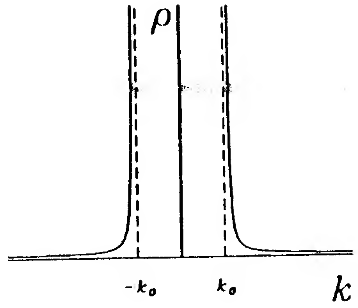

Figure 3: The weight accorded to a vacuum state $|\Omega_0;k\rangle$, with an infinite peak at k_0 where the effective cosmological constant for the state vanishes.

universe with an almost vanishing effective cosmological constant. The four-form vacuum selected is such that it cancels the vacuum energy of the rest of the particle physics. In general the particle physics vacuum is a superposition of the k-vacua,

$$|\Omega\rangle = \int dk\, c(k)|\Omega_0;k\rangle$$

and we have been led to the conclusion that, $|c(k)|^2 \sim \exp(3\pi/G_N\Lambda(k))$. Hence $|\Omega\rangle$ is saturated by the vacua possessing zero cosmological constant.

9 Wormholes

One of the most radical developments in QGD over the past few years has been the subject of wormholes. The idea is that the universe has many topologically disjoint regions which are connected by tunnelling configurations called wormholes[7]. Thus what we consider as our universe is only one of a vast ensemble of universes which interact with each other. The interaction is through their *global* properties. This has profound consequence [45], as we shall see, of making the fundamental constants in our universe 'variable' in the sense of section 7. The interaction with the ensemble gives rise to a distribution of possible values. It has been suggested that when both the wormholes and other universes are traced out of the theory

[7] The 'wormholes' in this subject are four-dimensional instantons, not to be confused with the Schwarzschild wormholes beloved of science fiction writers. For a discussion of the latter I refer you to MTW [4]. Instructions for building a time machine can be found in ref. [44].

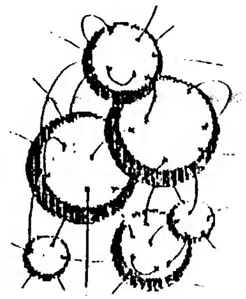

Figure 4: A collection of large deSitter universes in Euclidean space (four-spheres) connected by small wormholes.

that the cosmological constant in the resulting one-universe theory is driven to zero [46]. Moreover, it is claimed that other fundamental constants of nature could also have their values fixed by this mechanism.

There are three big assumptions that go into the wormhole calculus. The first is that it makes sense to talk about quantum gravity as a path integral over Euclidean metrics. The second is that Euclidean path integrals can be approximated, semi-classically, by a saddle-point evaluation about their stationary value. The third is that there is a sharp distinction between large scales (universes) and small scales (wormholes). We have already encountered the Euclideanized version of deSitter space as a four-sphere of radius $\sqrt{3/\Lambda}$. If the radius of the space is very large then at small distances it will appear to be almost flat. A wormhole is a Euclidean geometry which has two asymptotically flat regions connected by a throat of characteristic radius r_w. This has to be larger than the Planck length, otherwise the theory of gravity we are using breaks down. We assume $r_w \gtrsim M_P^{-1}$ so that the throat is much smaller than the radii of two universes and the asymptotic regions of the wormhole can then patch onto the spheres and connect them. Another possibility is that both ends of the wormhole patch onto the same sphere. In both situations it costs very little action to do this compared to the action for the universes. A typical configuration is sketched in Figure 4. A discussion of one type of wormhole, the axionic wormhole, is to be found in the appendix.

If we accept the third big assumption above we should be able to separate the path

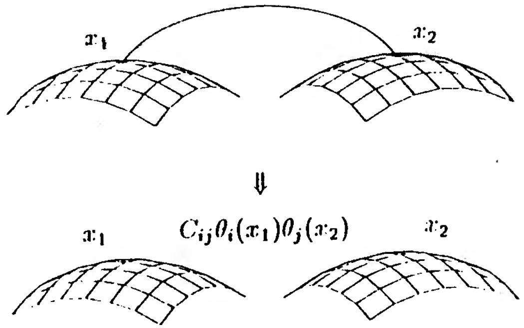

Figure 5: The effect of integrating out a wormhole connecting two points is to introduce an effective interaction of the form $C_{ij}\theta_i(x_1)\theta_j(x_2)$, where the $\theta_i(x)$ are local operators on the manifold.

integral of everything into wormholes and universes. The complete path integral, wormholes and all, has the form,

$$Z_E = \int \mathcal{D}g_{\mu\nu}\, \mathcal{D}\phi \, \exp\left(-S_E(g,\phi)\right) \tag{56}$$

and from it we can find the expectation values of operators as usual. We would like to know where our universe fits into all of this. Therefore we should integrate out all the wormholes. We would then be left with an effective theory, valid at length scales larger than the characteristic wormhole size. A wormhole can connect any two points, x_1 and x_2 say, of the manifold as shown in Figure 5. When this wormhole is integrated out of (56) what remains is an effective interaction between the degrees of freedom at the two points. We can represent this by an insertion of a factor into the path integral,

$$\int \mathcal{D}g_{\mu\nu} \longrightarrow \int \bar{\mathcal{D}}g_{\mu\nu} \sum_{ij} C_{ij}\theta_i(x_1)\theta_j(x_2)$$

where the integration is now over all geometries excluding the one wormhole. The $\theta_i(x)$ are some (gauge invariant) local operators of the theory, and the reason they are factorized is that it is assumed the two ends of the wormhole do not interact with each other (the dilute gas approximation) and they can stick onto anywhere they like. The C_{ij} encapsulate the unknown properties of the wormhole. When we integrate over all possible positions where

the wormhole ends can attach themselves we find that the insertion in this case is,

$$K = \frac{1}{2}\sum_{ij} C_{ij}\Theta_i\Theta_j$$

where the operators are $\Theta_i = \int d^4x\sqrt{g(x)}\,\theta_i(x)$, and the factor of a half is because both ends of the wormhole are identical and we have to avoid counting their attachments twice.

When we integrate out n wormholes the insertion would be $K^n/n!$ as there are $n!$ permutations of the identical wormholes. Summing over all possible wormhole configurations therefore gives us an insertion factor of,

$$F_w = \exp\left[\frac{1}{2}\sum_{ij} C_{ij}\Theta_i\Theta_j\right]$$

into the path integral. As we can see this is an interaction that is global in nature. This can be rewritten in the form of a gaussian integral from the relation,

$$F_w = N\int\left(\prod_i d\alpha_i\right)\exp\left[-\frac{1}{2}\sum_{ij} C_{ij}^{-1}\alpha_i\alpha_j\right]\exp\left[\sum_i \alpha_i\Theta_i\right] \tag{57}$$

The net effect of integrating all the wormholes out of the theory, tracing over their configurations, is the insertion of the factor F_w into the path integral. With the form of (57) we find that,

$$\operatorname*{Tr}_{\text{wormholes}}\left[\int \mathcal{D}g_{\mu\nu}\,\mathcal{D}\phi\, e^{-S_E}\right] = \int d\alpha P(\alpha)\left[\int \bar{\mathcal{D}}g_{\mu\nu}\,\mathcal{D}\phi\, e^{-S_E(\alpha)}\right] \tag{58}$$

where the path integral is now over all *large* geometries. This is depicted in Figure 6. We have absorbed the second term of F_w into the definition of an effective action,

$$S_E(\alpha) = S_E - \sum_i \alpha_i\Theta_i \tag{59}$$

Recall that the Θ_i are operators integrated over the spacetime. The original action is also expressible in these terms as $S_E = \sum_i \lambda_i\Theta_i$ where the λ_i are the strengths of these terms in the action *i.e.* they are essentially the coupling constants of the theory. The implication of (59) is that the constants of nature undergo a shift $\lambda_i \to \lambda_i - \alpha_i$, one for each coupling. This shift is linear because of our initial assumption that wormholes don't interact with each other. The probability distribution for these shift parameters is,

$$P(\alpha) = \exp\left[-\frac{1}{2}\sum_{ij} C_{ij}^{-1}\alpha_i\alpha_j\right] \tag{60}$$

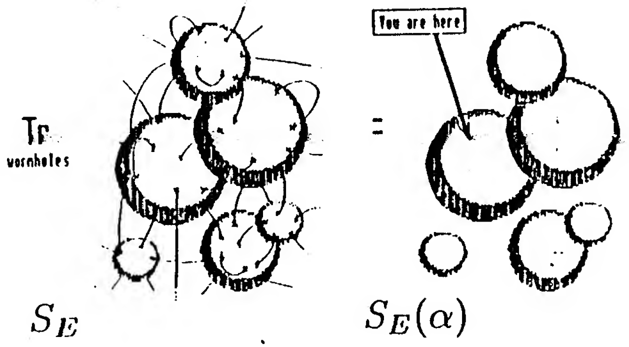

Figure 6: Tracing out the wormholes disconnects the manifolds and results in an effective theory for length scales greater than r_w.

We have therefore satisfied the first criterion of the BH mechanism, namely allowing the constants of nature to vary, their values determined by some external means.

We are not quite finished yet. In integrating out the wormholes we have severed the links between different points on large manifolds. These points need not be common to the same large manifold, and so we end up with topologically disjoint pieces of the original huge manifold. The path integral over large manifolds in (58) is really summing over disjoint large geometries. Again we want an effective theory for *one* universe and so we have to trace out all the others. Each topologically distinct region will contribute,

$$Z(\alpha) = \int \mathcal{D}g_{\mu\nu}\,\mathcal{D}\phi\, e^{-S_E(\alpha)} \tag{61}$$

to the path integral in the right hand side of (58). This is the path integral over all the geometrical configurations of just one universe. If we have n extra universes to contend with then the effect of tracing them out would be a factor $Z^n/n!$. Thus summing over all configurations gives a factor of e^Z in (58). We then arrive at the effective theory for one universe,

$$\mathop{\mathrm{Tr}}_{\substack{\text{wormholes}\\ \text{universes}}}\left[\int \mathcal{D}g_{\mu\nu}\,\mathcal{D}\phi\, e^{-S_E}\right] = \int d\alpha\, P(\alpha)Q(\alpha)\left[\int \mathcal{D}g_{\mu\nu}\,\mathcal{D}\phi\, e^{-S_E(\alpha)}\right] \tag{62}$$

where the function $Q(\alpha) = \exp(Z(\alpha))$ with $Z(\alpha)$ given by (61). We can calculate, for example, expectation values in the one universe theory using the path integral $Z(\alpha)$ and the values we get will depend on the external wormhole parameters α. Different values of

α correspond to different theories. The distribution of the parameters is determined by the functions $P(\alpha)$ and $Q(\alpha)$ (the *deus ex machina* again). The second part of the BH mechanism is to see if these pick out any special values for the couplings.

If we accept the second big assumption, that path integrals over Euclidean geometries can be approximated by a saddle point through their stationary solution, then we might be justified [46] in taking $Z(\alpha) \sim \exp(-\tilde{S}_E(\alpha))$. As we have seen in previous sections, the stationary value of the Euclidean action for a universe with cosmological constant Λ is $\tilde{S}_E = -3\pi/G_N\Lambda$. In the wormhole picture the constants of nature depend on the α and so we believe we can approximate,

$$Q(\alpha) \sim \exp\left(\exp\left(\frac{3\pi}{G_N(\alpha)\Lambda(\alpha)}\right)\right) \tag{63}$$

Indeed this was the result first obtained by Coleman. If we ignore the α dependence of Newton's constant for the moment we would conclude that of all the αs there is a subset of them for which $\Lambda(\alpha) \to 0_+$ since for this subset the weight function $Q(\alpha) \to \infty$. Thus the integral in (62) is dominated by the α for which the cosmological constant is driven to zero. The one universe theory has a vanishing cosmological constant from the effects of the wormholes.

However, there are problems with this picture as I shall now briefly summarize.

- Equation (63), if correct, also implies that Newton's constant is driven to zero [47, 48]. To avoid this we have to further assume that $G_N(\alpha)$ is bounded from below as a function of all the wormhole parameters. Then on the subset for which $\Lambda(\alpha) \to 0_+$ Newton's constant attains its minimum allowed on this subset.

- There are many types of wormholes with a spectrum of characteristic scales r_w. We have been assuming all along that we can integrate them out of (58) in one go. To a small wormhole a patch of a large wormhole looks much the same as a patch of a large sphere. Integrating out small wormholes might still leave some large wormholes in the sum over large geometries. We said that it costs very little action to insert a wormhole ($S_w \sim r_w^2 M_P^2$) and we would expect the coefficients to be $C_{ij} \sim e^{-S_w}$, implying a suppression of large wormhole effects. However, the e^{-S_w} can be absorbed into a redefinition of the α from (57) and so it appears that wormholes with domestic scales can indeed be generated [49].

- Perhaps the most damaging result is that (63) is using the wrong approximation to the path integral. Certainly we get a factor of $\exp(-\tilde{S}_E(\alpha))$ in the approximation to $Z(\alpha)$, but we have still neglected the phase of this coming from the determinant of the fluctuations around the saddle point (this corresponds to the phase of the steepest descent contour in standard complex analysis). Polchinski [50] has calculated this phase and found it to be π. With this minus sign, Coleman's result disappears.

Some other criticisms of the wormhole calculus are levelled at the use of Euclidean path integrals in the first place [51].

Nevertheless, if we just accept that minimizing the extremal action $\tilde{S}_E(\alpha)$ as a function of the α fixes the constants of nature to special values we might be able to calculate them. This is what is commonly called "the big fix". The results show that scalar masses have a tendency to be much smaller than the wormhole scale [47], while fermions not protected by chiral symmetries tend to have their masses driven to higher values. However, in an interacting theory there is a compromise between these opposing effects and it is possible to get an indication of how the masses and couplings of the standard model, for example, might be constrained [52].

Another constant of nature which has been investigated in this context is the strong CP violation parameter θ_{QCD}. The manner in which θ_{QCD} contributes to physical processes is purely through nonperturbative strong interactions, characterized by a scale Λ_{QCD}. Its effects are irrelevant at high energies due to asymptotic freedom. The Yukawa couplings of the quarks to the Higgs boson are dependent on the α parameters. These transmute to the quark mass matrix when electroweak symmetry is spontaneously broken. Upon diagonalization θ_{QCD} picks up the phase of this matrix through the colour anomaly and therefore acquires a dependence on α. On the surface $\Lambda(\alpha) = 0$ in α space Newton's constant attains its minimum by the arguments above. However, G_N depends on θ_{QCD} through the α and, considered as a function of θ_{QCD}, it must be minimized. The important point is that the dependence of G_N on θ_{QCD} can be calculated solely from low energy physics (pion loops). Chiral perturbation arguments [53] show that $G_N(\theta_{QCD})$ is minimized when $m_\pi^2(\theta_{QCD})$ is. In the chiral limit (where the quark masses are small) we know [54] that for two quarks,

$$m_\pi^2(\theta_{QCD}) = \frac{v}{f_\pi^2}\left(m_u^2 + m_d^2 + 2m_u m_d \cos\theta_{QCD}\right)^{1/2}$$

where f_π is the pion decay constant and $v = \left|\langle 0|\bar{u}u + \bar{d}d|0\rangle\right|$. The implication is that wormhole effects drive θ_{QCD} to one of its CP conserving values of π, there to be slightly corrected by electroweak effects ($\Delta\theta_{QCD} \lesssim 10^{-9}$ experimentally). The fact that the value of $\theta_{QCD} \sim \pi$ is the one preferred by wormholes is at odds with the usual prediction of $\theta_{QCD} \sim 0$ via the Peccei-Quinn mechanism. However, it is still uncertain what the experimentally preferred value is. Once this is settled we at least have the chance of either confirming the prediction or refuting the theory.

10 Concluding remarks

The ideas of QGD may seem unusual on first encounter (Einstein, with his reluctance to accept the implications of quantum mechanics, would have probably regarded its use here a perversion). However, further consideration leads us to believe that it might be a good effective description of spacetime as we approach the Planck scale. Some of the results are encouraging. QGD can provide a framework for discussing questions that would previously have seemed metaphysical. It does so not by positing a grand unified theory of everything, but rather by taking what we do know and reformulating it in another language. We have seen that in doing so it imposes constraints on our particle and cosmological models. This in turn leads to predictions for the physics that we are accustomed to, and so QGD can in principle be refuted by experiment. As we have seen, this approach to quantum gravity also raises some interesting new questions. I hope the student will find time to explore them further.

A The Axionic wormhole

It was the discovery of the axionic wormhole that inspired the development of the wormhole program. Giddings and Strominger ~~[45] were the first to note that a third rank antisymmetric~~ tensor (this is equivalent to an axion as we shall see) coupled to gravity can have instanton solutions. These instantons, wormholes, describe the quantum tunnelling between spacetimes of different topology. Since then many other classes of wormhole solutions have been found (for a partial list see ref. [55, 56]). However, we shall restrict our discussion to that of the axionic wormhole as it is easy to construct [55, 57, 25] and has simple properties. In what follows we shall first show how a three-form field is equivalent to an axion, and then explicitly construct the tunnelling solution.

A three-form field strength appears quite naturally [58] in superstring theories. From considerations of the behaviour of the theory in ten dimensions we know that the field strength must take the form $\mathbf{H} = \mathbf{dB} + \omega_L - \omega_Y$. Here $\mathbf{B}$ is a two-form potential (and is part of a larger supergravity multiplet in string theory). The two corrections are the Lorentz and Yang-Mills Chern-Simons terms respectively, necessary for the gauge invariance. The same relationship holds between the components that fully reside in our familiar world of four dimensions. We shall now just assume this expression for $\mathbf{H}$ in four dimensions without bothering any further where it comes from. The field strength is globally defined, and we can then construct an action,

$$S_H = -\frac{1}{2f^2}\int_{\mathcal{M}} \mathbf{H}\wedge *\mathbf{H} = -\frac{1}{2f^2}\frac{1}{3!}\int_{\mathcal{M}} H_{\mu\nu\lambda}H^{\mu\nu\lambda}\sqrt{-g}\,d^4x \tag{64}$$

similar to the Maxwell action. Here f is a parameter with the dimensions of mass. By varying this with respect to the potential $B_{\mu\nu}$ we find that the resulting equation of motion is simply $\mathbf{d} * \mathbf{H} = 0$. The solution of this is that the dual of $\mathbf{H}$ is exact *i.e.* $\mathbf{H} = f * \mathbf{d}\varphi$ (this is reminiscent of the four-form case considered before). We see that the field φ is a pseudoscalar due to its definition through duality, $H_{\mu\nu\lambda} = f\epsilon^{\rho}{}_{\mu\nu\lambda}(\partial_\rho\varphi)$.

It is easy to demonstrate that this pseudoscalar field is an axion. By the definition of the three-form above we find that its Bianchi identity is $\mathbf{dH} = \mathrm{Tr}(\mathbf{R}\wedge\mathbf{R}) - \mathrm{Tr}(\mathbf{F}\wedge\mathbf{F})$. Expressed in terms of the pseudoscalar this Bianchi identity is,

$$f\Box\varphi = \frac{1}{2}\mathrm{Tr}(F_{\mu\nu}\tilde{F}^{\mu\nu}) - \frac{1}{2}\mathrm{Tr}(R_{\mu\nu}\tilde{R}^{\mu\nu}) \tag{65}$$

where $\tilde{F}_{\mu\nu} = \frac{1}{2}\epsilon_{\mu\nu}{}^{\lambda\rho}F_{\lambda\rho}$ and $\tilde{R}^{\alpha}{}_{\beta\mu\nu} = \frac{1}{2}\epsilon_{\mu\nu}{}^{\lambda\rho}R^{\alpha}{}_{\beta\lambda\rho}$ are the dual of the gauge field strength and Riemann tensors respectively. This is the equation of motion for an axion. Similarly the stress tensor for the three-form is found, as usual, by variation of (64) with respect to the ~~metric~~,

$$T_{\mu\nu} = \frac{1}{3!f^2}\left(3H_{\mu\lambda\rho}H_{\nu}{}^{\lambda\rho} - \frac{1}{2}g_{\mu\nu}H_{\kappa\lambda\rho}H^{\kappa\lambda\rho}\right)$$

When re-expressed in terms of the pseudoscalar it is seen to be the stress tensor for a massless scalar field. For a review of axions I refer you to the TASI lectures of Kim [54]. Incidentally, the non-minimal coupling to gravity in (65) can give rise to dipole hair (and therefore no axionic charge) on a black hole [59]. In the case of minimal coupling ($\mathbf{dH} = 0$) the black hole would carry an axionic charge [60] but nevertheless has no hair.

Instantons are solutions to the Euclidean version of a theory and which interpolate between certain initial and final Lorentzian configurations. Let us just construct the wormhole first and discuss its nature later. The starting point is to find solutions to the Euclideanized equations of motion. The Euclidean action for the three-form coupled to gravity is,

$$S_E = -\frac{1}{16\pi G_N}\int_{\mathcal{M}} R\sqrt{g}\,d^4x - \frac{1}{8\pi G_N}\int_{\partial\mathcal{M}}\left(K - K^0\right)\sqrt{h}\,d^3x + \frac{1}{2f^2}\int_{\mathcal{M}}\mathbf{H}\wedge *\mathbf{H} \tag{66}$$

If the manifold has a boundary we add the surface contribution with the extrinsic curvature as we did previously. However, we have to correct this if the manifold is asymptotically flat by subtracting the extrinsic curvature of the boundary embedded in a flat four-space, K^0_{ij}. Why we have to do this will be explained shortly.

We take the spatial slices of this manifold to be three-spheres as before. The metric is therefore the Euclideanized form of (13),

$$ds^2 = \left(\frac{2G_N}{3\pi}\right)\left(N_E^2(\tau)\,\mathbf{d}\tau\otimes\mathbf{d}\tau + a^2(\tau)\,\delta_{ij}\,\sigma^i\otimes\sigma^j\right) \tag{67}$$

With this metric the trace of the extrinsic curvature is $K = \left(\frac{2G_N}{3\pi}\right)^{-1/2}\left(\frac{3\dot{a}}{N_E a}\right)$. The metric for a flat four-space with the same value for the radius of the three-sphere is (67) with $N_E = |\dot{a}|$, and thus the trace of the extrinsic curvature embedded in this flat space is $K^0 = \left(\frac{2G_N}{3\pi}\right)^{-1/2}\left(\frac{3}{a}\right)\mathrm{sign}(\dot{a})$. It can also be shown [59] that for homogeneous and isotropic slices, such as the above, $\mathrm{Tr}(\mathbf{R}\wedge\mathbf{R}) = 0$. Thus if we have trivial gauge fields the Bianchi identity for the three-form field strength is $\mathbf{dH} = 0$.

The field strength should be homogeneous and isotropic. Therefore we are able to express it in terms of the basis for the three-spheres, $\mathbf{H} = \left(\frac{2G_N}{3\pi}\right)^{1/2} h(\tau)\,\sigma^1\wedge\sigma^2\wedge\sigma^3$. The Bianchi identity then tells us that $h(\tau)$ is a constant of the motion. In fact if we consider a piece of the manifold bounded by two spatial sections $\partial\mathcal{M}_\pm$ then,

$$\int_{\partial\mathcal{M}_+} \mathbf{H} = \int_{\partial\mathcal{M}_-} \mathbf{H} = (2\pi^2)\left(\frac{2G_N}{3\pi}\right)^{1/2} h \tag{68}$$

This reflects the conservation of an axionic charge. The action for the pseudoscalar is invariant under a constant shift, $\varphi \to \varphi + c$, and you can check that the conserved charge for this is just the above.

Re-expressing (66) in terms of the above and performing the integrals over the homogeneous spatial slices we find that it has the form,

$$S_E = \frac{1}{2}\int\left(-\frac{\dot{a}^2 a}{N_E} - N_E a + \frac{h^2 N_E}{f^2 a^3}\right)d\tau + \frac{1}{2}\left[a^2 \mathrm{sign}(\dot{a})\right]_{\partial\mathcal{M}} \tag{69}$$

The Hamiltonian constraint from the above yields the following equation of motion for the scale factor a (in the 'proper time' gauge $N_E = 1$),

$$\dot{a}^2 = 1 - \frac{\hat{r}_w^4}{a^4} \tag{70}$$

where $\hat{r}_w^4 = h^2/f^2$. This can be solved in terms of elliptic integrals [42]. However the properties of this four-space should be clear from inspection of (70). As $a \to \infty$ then $\dot{a} \to \pm 1$ meaning that the geometry is asymptotically flat Euclidean space. Moreover the scale factor attains its minimum when $a = \hat{r}_w$ in Planck units. This is sketched in Figure 7. This wormhole can then connect two flat Euclidean spaces. It carries a charge given by (68).

The value of the Euclidean action for this is readily calculated from (69). If we integrate the classical solution (70) between the slices on either side of the throat for which $a = R$ the action for this can be written,

$$S_E(R) = -2\int_{\hat{r}_w}^{R} a\sqrt{1 - \frac{\hat{r}_w^4}{a^4}}\,da + R^2$$

In the limit of $R \to \infty$ this gives the action for the wormhole $S_w = \frac{\pi}{2}\hat{r}_w^2$, or converting back from the Planck units $S_w = \frac{3\pi^2}{4} r_w^2 M_P^2$. You can now see the rationalle for subtracting the boundary contribution embedded in flat space in (66). The wormhole connects flat spaces

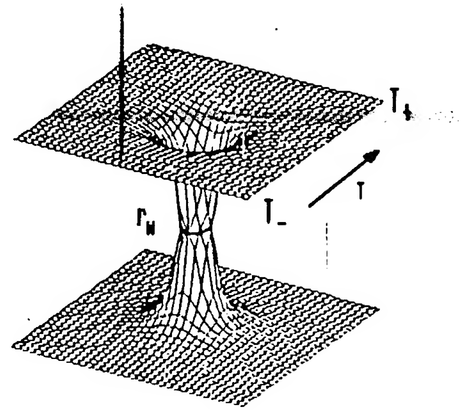

Figure 7: The wormhole solution with asymptotically flat Euclidean sections and whose throat is a three-sphere of radius r_w.

and so we really are interested in the difference between the actions for the two flat spaces with and without the wormhole. Thus we had to remove the flat space actions. If these had been present the action would have diverged.

We now come to the Minkowski space interpretation of this as an instanton for topology change. In Figure 7 the rectangular grid far from the mouth is intentional. We could equally well have decided to draw the picture as a series of concentric rings representing the three-spheres emerging from the mouths with the 'axial' direction down the throat repesenting the τ values. These would be the coordinates we employed in (67). In the asymptotic regions these rings would have been evenly spaced apart since $\dot{a} \to \pm 1$. Transforming to the rectangular coordinates we can then call one of the directions T say. Consider the two lines $T_\pm$ shown in the figure. If these are far enough away from the mouth the extrinsic curvature along these constant T slices is vanishingly small. If we now take a flat Minkowski spacetime and consider a cut at some time value t then we can identify the spatial slice just before t with the spatial slice T_-, and the one just after t with T_+. We can do the same with the other mouth and another piece of Minkowski space. This is the instanton. There is an instantaneous change in the topology and the wormhole carries off some axionic charge. Equally well we could cut at the throat. The extrinsic curvature also vanishes at the throat and we could similarly identify this three-surface with that of a closed FRW universe at its point of maximal expansion. This represents the creation of a 'baby universe'. There is a

discontinuous jump in Minkowski space and the appearance of a FRW universe which carries the axion charge lost by the flat space.

References

[1] A.Strominger, in *Particles, Strings an Supernovae. Proceedings of TASI 1988*, eds. A.Jevicki and C.-I.Tan (World Scientific, Teaneck NJ, 1989);
I.Klebanov, in *From Actions to Answers. Proceedings of TASI 1989*, eds. T.DeGrand and D.Toussiant (World Scientific, Teaneck NJ, 1990).

[2] S.W.Hawking in *Relativity, Groups and Topology II*, eds. B.S.DeWitt and R.Stora (Elsevier Science Publishers, 1984);
Quantum Cosmology, eds. L.Z.Fang and R.Ruffini (World Scientific, Singapore, 1987).

[3] J.J.Halliwell, Int. J. Mod. Phys. **A5** (1990) 2473.

[4] S. Weinberg, *Gravitation and Cosmology* (John Wiley & Sons, New York, 1972);
C.W.Misner, K.S.Thorne and J.A.Wheeler, *Gravitation* (Freeman, San Fransisco, 1973);
L.Landau and E.Lifshitz, *The Classical Theory of Fields* (Pergamon Press, 1975);
R.M.Wald, *General Relativity* (University of Chicago Press, 1984).

[5] R.Arnowitt, S.Deser and C.W.Misner, in *Gravitation, an Introduction to Current Research*, ed. L.Witten (Wiley, New York, 1962)

[6] J.M.York, Phys. Rev. Lett. **28** (1972) 1082;
G.W.Gibbons and S.W.Hawking, Phys. Rev. **D15** (1977) 2752.

[7] B.S.DeWitt, Phys. Rev. **160** (1967) 1113;
J.A.Wheeler, in *Battelle Recontres*, eds. C.DeWitt and J.A.Wheeler (Benjamin, New York, 1968);
C.W.Misner, Phys. Rev. **186** (1969) 1319.

[8] T.Eguchi, P.B.Gilkey and A.J.Hanson, Phys. Rep. **66** (1980) 213.

[9] S.W.Hawking and D.N.Page, Nucl. Phys. **B264** (1986) 185.

[10] C.Teitelboim, Phys. Rev. **D25** (1982) 3159;
K.Kuchař, J. Math. Phys. **24** (1983) 2122;
J.J.Halliwell, Phys. Rev. **D38** (1988) 2468.

[11] A.H. Guth, Phys. Rev. **D23** (1981) 347;
A.D. Linde, Phys. Lett. **108B** (1982) 389;
A. Albrecht and P.J. Steinhardt, Phys. Rev. Lett. **48** (1982) 1220;
A.D. Linde, Phys. Lett. **129B** (1983) 177;
The following three references contain reviews on Inflation:
M.S. Turner, in *Proceedings of the Cargèse School on Fundamental Physics and Cosmology*, eds. J. Audouze and J. Tran Thau Van (Editions Frontières, 1985);
R.H. Brandenberger, Rev. Mod. Phys. **57** (1985) 1;
K.A. Olive, Phys. Rep. **190** (1990) 307.

[12] V.G.Lapchinsky and V.A.Rubakov, Acta. Phys. Polonica **B10** (1979) 1041;
T.Banks, W.Fischler and L.Susskind, Nucl. Phys. **B262** (1985) 159;
J.J.Halliwell and S.W.Hawking, Phys. Rev. **D31** (1985) 1777;
R.Brout and G.Venturi, Phys. Rev. **D39** (1989) 2436;
J.J.Halliwell, Santa Barbara preprint NSF-ITP-88-126 (1988).

[13] M.Born and J.R.Oppenheimer, Ann. der Phys. **84** (1927) 457;
H.A.Bethe and R.Jackiw, *Intermediate Quantum Mechanics* (Benjamin, Menlo Park CA, 1986).

[14] H.Goldstein, *Classical Mechanics* (Addison-Wesley, Reading MA, 1980).

[15] M.V.Berry, Proc. Roy. Soc. London Ser. **A392** (1984) 45.

[16] N.D.Birrell and P.C.W.Davies, *Quantum Fields in Curved Space* (Cambridge University Press, 1982).

[17] B.Spinoza, *Ethics* (Dent, London, 1989).

[18] J.J.Halliwell, Phys. Rev. **D36** (1987) 3626;
A.Vilenkin, Phys. Rev. **D39** (1989) 1116;

J.J.Halliwell, Phys. Rev. **D39** (1989) 2912;
M.Morikawa, Phys. Rev. **D42** (1990) 2929.

[19] A.D.Linde, Nuovo. Cim. Lett. **39** (1984) 401;
A.D.Linde, in *Proceedings of the 1983 Shelter Island Conference on Quantum Field Theory and Fundamental Problems of Physics*, eds. R.Jackiw, N.Khuri, S.Weinberg and E.Witten (MIT Press, 1985);
I.G.Moss and W.A.Wright, Phys. Rev. **D29** (1984) 1067.

[20] A.D.Linde, *Particle Physics and Inflationary Cosmology* (Harwood Academic, Switzerland, 1990).

[21] *The Many-Worlds Interpretation of Quantum Mechanics*, eds. B.S.DeWitt and N.Graham (Princeton University Press, 1973);
J.S.Bell, *Speakable and Unspeakable in Quantum Mechanics* (Cambridge University Press, 1987);
J.D.Barrow and F.J.Tipler, *The Anthropic Cosmological Principle* (Oxford University Press, 1988).

[22] J.B.Hartle and S.W.Hawking, Phys. Rev. **D28** (1983) 2960.

[23] Ya.B.Zeldovich and A.A.Starobinsky, Sov. Astron. Lett. **10** (1984) 135;
A.D.Linde, JETP **60** (1984) 211, Lett. Nuovo Cim. **39** (1984) 401;
V.A.Rubakov, Phys. Lett. **148B** (1984) 280;
A.Vilenkin, Phys. Rev. **D30** (1984) 509.

[24] *Quantum Theory and Measurement*, eds. J.A.Wheeler and W.H.Zurek (Princeton University Press, 1983).

[25] M.J.Duncan and L.G.Jensen, Nucl. Phys. **B328** (1989) 171.

[26] R.P.Feynman and A.R.Hibbs, *Quantum Mechanics and Path Integrals* (McGraw-Hill, New York, 1965);
E.S.Abers and B.W.Lee, Phys. Rep. **9** (1973) 1;
P.Ramond, *Field Theory, A Modern Primer*, (Benjamin, Reading, 1981);

L.S.Schulman, *Techniques and Applications of Path Integrals* (Wiley and Sons, New York, 1981).

[27] E.Farhi, Phys. Lett. **219B** (1989) 403.

[28] C.Teitelboim, Phys. Rev. Lett. **50** (1983) 705.

[29] D.W.McLaughlin, J. Math. Phys. **13** (1972) 1099.

[30] J.J.Halliwell and J.Luoko, Phys. Rev. **D39** (1989) 2206, **D40** (1989) 1868, **D42** (1990) 3997;
J.J.Halliwell and J.B.Hartle, Phys. Rev. **D41** (1990) 1815;
J.D.Brown and E.A.Martinez, Phys. Rev. **D42** 1931;
C.Kiefer, Zurich preprint 90-0238 (1990).

[31] S.Weinberg, Rev. Mod. Phys. **61** (1989) 1.

[32] E.Baum, Phys. Lett. **133B** (1983) 185.

[33] S.W.Hawking, Phys. Lett. **134B** (1984) 403.

[34] M.J.Duff, Phys. Lett. **226B** (1989) 36.

[35] M.J.Duncan and L.G.Jensen, Nucl. Phys. **B336** (1990) 100.

[36] A.Aurelia, H.Nicolai and P.Townsend, Nucl. Phys. **B176** (1980) 509;
M.J.Duff and P.van Nieuwenhuizen, Phys. Lett. **94B** (1980) 179;
P.G.O.Freund and M.Rubin, Phys. Lett. **97B** (1980) 233.

[37] M.Henneaux and C.Teitelboim, Phys. Lett. **143B** (1984) 415;
E.Witten, in *Proceedings of the 1983 Shelter Island Conference on Quantum Field Theory and Fundamental Problems of Physics*, eds. R.Jackiw, N.Khuri, S.Weinberg and E.Witten (MIT Press, 1985);
J.D.Brown and C.Teitelboim, Phys. Lett. **195B** (1987) 177.

[38] V.A.Rubakov, Phys. Lett. **214B** (1988) 503.

[39] M.McGuigan, Phys. Rev. **D38** (1988) 3031;
A.~~Hosoya and~~ M.Morikawa, Phys. Rev. **D39** (1989) 1123;
M.McGuigan, Phys. Rev. **D39** (1989) 2229.

[40] S.B.Giddings and A.Strominger, Nucl. Phys. **B321** (1989) 481;
W.Fischler, I.Klebanov, J.Polchinski and L.Susskind, Nucl. Phys. **B327** (1989) 157.

[41] M.J.Duncan and L.G.Jensen, Minnesota preprint UMN-TH-908/90.

[42] M.Abramowitz and I.A.Stegun, *Handbook of Mathematical Functions* (Dover, New York, 1972).

[43] N.N.Bogolubov, Sov. Phys. JETP, **7** (1958) 51.

[44] M.S.Morris, K.S.Thorne and U.Yurtsever, Phys. Rev. Lett. **61** (1988) 1466.

[45] S.W.Hawking, Phys. Lett. **195B** (1987) 337;
S.W.Hawking, Phys. Rev. **D37** (1988) 904;
S.B.Giddings and A.Strominger, Nucl. Phys. **B306** (1988) 890;
S.Coleman, Nucl. Phys. **B307** (1988) 867.

[46] S.Coleman, Nucl. Phys. **B310** (1988) 643;
I.Klebanov, T.Banks and L.Susskind, Nucl. Phys. **B317** (1989) 665.

[47] B.Grinstein and M.B.Wise, Phys. Lett. **212B** (1988) 407.

[48] J.Preskill, Nucl. Phys. **B323** (1989) 141.

[49] W.Fischler and L.Susskind, Phys. Lett. **217B** (1989) 48;
S.Coleman and K.Lee, Phys. Lett. **221B** (1989) 242;
J.Polchinski, Nucl. Phys. **B325** (1989) 619.

[50] J.Polchinski, Phys. Lett. **210B** (1989) 251.

[51] W.G.Unruh, Phys. Rev. **D40** (1989) 1053;
G.Veneziano, Mod. Phys. Lett. **A4** (1989) 695.

[52] B.Grinstein and C.T.Hill, Phys. Lett. **B220** (1989) 520;
J.Ellis, E.Floratos and D.V.Nanopoulos, Phys. Lett. **B225** (1989) 121.

[53] H.B.Nielsen and M.Ninomiya, Phys. Rev. Lett. **62** (1989) 1429;
J.Preskill, S.P.Trivedi and M.B.Wise, Phys. Lett. **B223** (1989) 26;
K.Choi and R.Holman, Phys. Rev. Lett. **62** (1989) 2575;
K.Choi and R.Holman, Phys. Rev. Lett. **64** (1990) 131.

[54] J.E.Kim, in *Particles, Strings an Supernovae. Proceedings of TASI 1988*, eds. A.Jevicki and C.-I.Tan (World Scientific, Teaneck NJ, 1989).

[55] K.Lee, Phys. Rev. Lett. **61** (1988) 263.

[56] J.J.Halliwell and R.Laflamme, Class. Quant. Grav. **6** (1989) 1839;
L.F.Abbott and M.B.Wise, Nucl. Phys. **B325** (1989) 687;
S.Coleman and K.Lee. Nucl. Phys. **B329** (1990) 387;
L.F.Abbott and E.Farhi, Nucl. Phys. **B331** (1990) 263;
F.Accetta, A.Chodos and B.Shao, Nucl. Phys. **B333** (1990) 221.

[57] C.Burgess and A.Kshirsagar, Nucl. Phys. **B324** (1989) 157;
J.D.Brown *et. al.*, Nucl. Phys. **B328** (1989) 213.

[58] E.Witten, Phys. Lett. **149B** (1984) 351;
R.Rohm and E.Witten, Ann. Phys. **170** (1986) 454;
M.B.Green, J.H.Schwarz and E.Witten, *Superstring Theory* (Cambridge University Press, 1987).

[59] B.A.Campbell, M.J.Duncan, N.Kaloper and K.A.Olive, LAPP-TH-293-90 (Nucl. Phys. in press);
B.A.Campbell, M.J.Duncan, N.Kaloper and K.A.Olive, Phys. Lett. **251B** (1990) 34.

[60] M.J.Bowick *et. al.*, Phys. Rev. Lett. **61** (1988) 2823;
T.J.Allen, M.J.Bowick and A.Lahiri, Phys. Lett. **237B** (1990) 47.